DATE DUE

For each day's delay after the due date a fine of 3 P. per Vol. shall be charged for the first week, and 25 P. per Vol. per day for subsequent days.

Borrower's No.	Date Due	Borrower's No.	Date Due
871(1)	11.9.71.	2	

STATISTICS

By

Y. D. KESKAR V.D.

M. A., B. Com. F. S. S.

Late Principal, Sir Kikabhai Premchand
College of Commerce,
Surat.

DELHI POLYTECHNIC
LIBRARY
DELHI

VORA & CO., PUBLISHERS LIMITED
3, Round Building, Kalbadevi Road, Bombay 2.

Price Rs. Ten & As. Eight Only

Price Rs. Ten & As. Eight Only

DEDICATED TO THE MEMORY

OF

MY AFFECTIONATE WIFE

Mrs. MATHUBAI

PREFACE

The modern science of statistics no longer synonymous with 'Political Arithmetic' or is limited to Sociology, Political Economy or Demography. As Dr. Bowley has stated '*it is not merely a branch of political economy nor is it confined to any science*'. Its importance is being increasingly recognised in all spheres of life by economists, scientists, administrators as well as by business magnates. These days, hardly there is any office of importance—business or government—where a statistician does not play a role of importance. Realising this, universities too have prescribed courses in statistics in their various faculties.

As a teacher of statistics, I have noticed what a bugbear this easy subject has become for students mainly because of the absence of suitable books of elementary character and particularly non-mathematical in treatment. It is no wonder that many a student has to rely on cramming, the dead weight to reasonable understanding and that he forgets it no sooner he leaves the examination hall. This book has been specially designed with a view to create interest in young men and to assist them in a more genaral use of statistical methods in various spheres of life. Effort has been made to present this subject in a simple but comprehensive manner and to impart adequate grounding in the procedure for proper compilation, presentation and interpretation of statistical data. As a majority of students lack knowledge of higher mathematics, subjects like probabilits, regression, sampling, etc. have not been fully discussed in this elementary treatise.

A large number of diagrams and figures has been given to make the subject easy as well as interesting. Graded exercises and questions from various university papers have been given at the end of each chapter. In appendices important formulas, mathematical tables as well as a list of publications have been given for ready reference. In short, great care has been taken to make the book as useful as possible.

No doubt, this book on statistical method has been primarily meant for university students but it will be found helpful by businessmen also as in these days of competition and small margin of profit, old-fashioned methods must be scrapped and improved methods used for the expansion of business and for the increase of profits. The Dominion Statitician of Canada has rightly said. 'for this the statistics are the corner-stone-the basic organization, without which the endeavour to meet new situations will be seriously handicapped '.

In the end, the author has great pleasure in acknowledging his indebtedness to the various authors as well as friends and will deem it a favour if he is helped with suggestions as well as criticisms for its improvement. Mr. M. K. Vora deserves special thanks as without his urge probably this book would not have been completed.

Y. D. Keskar

Surat, 22nd July, 1948.

PREFACE TO THE SECOND EDITION

In bringing out this second edition effort has been made to incorporate all the suggestions received from various quarters At some places the matter has been re-arranged and also re-cast to make it more easy and intelligible. Wherever necessary the diagrams too have been changed. It is hoped that with these changes the book will be found more useful.

Surat
20th June, 1950

Y. D. K.

CONTENTS

Chapter Page

I. INTRODUCTORY 5

Historical Survey—Meaning and Scope of Statistics. Characteristics. Importance and Value. Limitations. Fallacies. Distrust of Statistics—Main Division of Statistics. Statistical Method and Applied Statistics—Role of a Statistician.

II STATISTICAL INVESTIGATION 17

Statistical investigation in the Ancient times. Stages of statistical investigation. Enumeration, Tabulation, Summarization and critical examination Preliminaries for the collection of data—Defining the problem-Planning. Purpose and Scope. Degree of Accuracy. Unit of statistical data. Types of Statistical units. Essential characteristics of statistical data

III. SAMPLING AND APPROXIMATION 25

Distinction between statistical description and statistical induction. Sampling. Census and Sampling. Law of Statistical Regularity. Law of Inertia of Large Numbers. Attributes of a good sample. Methods of sampling. Extensive sampling. Representative sampling. Random sampling. Stratified sampling. Reliability of a sample.

Approximation—its utility in statistics. Degree of accuracy. Methods of Approximation. Errors Absolute and relative. Sources of statistical errors. Biassed and unbiassed errors.

IV. COLLECTION OF STATISTICAL DATA 35

Primary and Secondary Data. Methods of collecting data—primary and secondary. Direct personal investigation. Indirect oral investigation. Estimates from correspondents. Schedules filled by informants. Schedules in charge of enumerators. Enumerators. Qualities necessary for them. Questionnaires. Essential requisites. Secondary data. Scrutiny of Secondary data. Important sources of secondary data. Choice of method.

V. CLASSIFICATION AND TABULATION. 49

Classification—Main object. Statistical data. Bases of Classification. Simple and manifold classification. Series and groups. Seriation. Continuous and discrete series. Tabulation. Stages in tabulation. Working Sheets. Crossing Fives. Crossing Tens. Spot diagram. Primary and derivative. Simple and complex tables. Manifold tabulation. Arrangement of tabular facts. Frequency tables. Variates. Class interval. Cumulative frequency table. More than and Less than tables. Advantages. Rules for Tabulation. Mechanical aids in tabulation.

VI AVERAGES 70

Averages. Differant forms of Averages. Arithmetic averages. Its Computation. Short-cut method. Composite average. Advantages and disadvantages.

Median—Meaning. Its Computation. Formula method. Graphic method. Advantages and disadvantages. Quartiles Deciles. Percentiles.

Mode—Meaning. Its computation. Different methods. Uses. Advantages and disadvantages.

Geometric Mean. Harmonic Mean. Relation between different averages. Essential requisites of an ideal average.

VII. DIAGRAMS. 112

Diagrams. Utility. Essential requisites. Various types Directions. Lines or Bars. Polar diagram. Area diagrams. Circles. Angular diagrams. Volume diagrams. Pictograms. Maps and Pictures. Isotype method. Pin maps.

VIII. GRAPHS. 145

Graphs. Utility. Construction. Rules to be observed. False base line. Frequency distribution graphs. Histogram. Frequency polygon. Frequency curve. More than and Less than diagrams. Time series graphs. Ratio scale graphs. Advantages and disadvantages. Comparison of the two types. Statistical charts. Galton's graph for locating median. Zee Chart. Zone and Band curves.

IX. HISTORICAL OR TIME SERIES. 183

Historical or Time series. Simple and cumulated series. Different types of movements. Basic or Long time trend Undulatory or Cyclical Movements. Seasonal variation. Irregular fluctuation. Measuring time Changes. S-curve. Method adopted to ascertain secular trend. Free-hand smoothing. A series of Averages. Moving Average. Progessive Average. Method of Least squares. Measurement of other fluctuations. Methods adopted. Monthly Average

method. Monthly averages corrected for trend. Link relative method. Moving Averages method.

X. DISPERSION. 209

Dispersion. Meaning. Absolute and relative dispersion. Measure of dispersion Range. Cumulative range Quartile deviation. Mean deviation. Standard deviation. Variance. Relation between different measures of dispersion. Coefficient of variation. Lorenz curve. Moments of dispersion. Modulus.

XI. SKEWNESS. 244

Skewness. Meaning. Dispersion and skewness. Measures of skewness. Position of Average Method. Quartile deviation method. Cubed deviation method. Coefficient of Skewness.

XII. CORRELATION. 256

Correlation. Meaning. Subject and Relative. Positive and Negative correlation. Degree of correlation. Lag and Lead. Methods of ascertaining correlation. Graphic Methods. Plotting the indices. Scatter or not diagram. Person's Method Interpretation of the coefficient. Correlation of Long and Short time series. Correlation of grouped data. Concurrent deviation method. Spearman's method of Ranks. Bracket rank method Mid-rank method. Spearman's foot-rule. Ratio of variation. Galton graph. Regression line.

XIII. INTERPOLATION AND EXTRAPOLATION. 302

Interpolation Meaning. Utility. Assumptions. Methods adopted. Graphic method. Algebric method. Fitting with the parabolic curve. Lagrange's formula

Extrapolation Meaning. Utility. Graphic method.

XIV. INDEX NUMBER. 319

Index Number. Meaning. Construction. Points to be considered. Purpose. Selection of items. Prices. Choice of Base. Form of average to be used. Methods of Aggregates. Relative of Aggregates method. Chain base method Uses. Various type of index numbers. Index of Business Activity in India. Working-Class Cost of Living Index Numbers for Bombay City. Economic Adviser's Index Numbers. All India Index Numbers of Wholesale Prices of Food Articles. Cost of Living Index Number (British Ministry of Labour). The Economist Index Number of Business Activity. Investment Index Number. Investor's Chronicle. Index of Cost of Living (U. S. Bureau of Labour). Dun's Index Number.

XV. INTERPRETATION OF STATISTICAL DATA. ... 343

Interpretation. Its technical nature. Difficulties. Prerequisites of interpretation. Adequacy, Stability, homogeniety, relevancy, accuracy, Scientific analysis. Errors in interpretation. General directions.

XVI. STATISTICAL MATERIAL IN INDIA. ... ...

Statistics in India in the past. The Indian Economic Enquiry Committee. Bowley—Robertson Committee Report. Organization of Statistics. Measurement of National Income. Rural surveys. Urban surveys. University city surveys. Census of production. Government publications. Source of statistics in India. Defects of Government statistics. Government statistical publication. Guide to Current Official Statistics. Agricultural Statistics. Transport and Communication statistics. Industrial statistics. Price statistics. Cost of Living statistics. Wages Statistics. Trade statistics. National Income. Census Reports. Intercensual estimates of population. Vital statistics. 358

APPENDIX I. Specimen of blank forms and questionnaires. ... 393

II. Logarithms and their uses. ... 423

III. Mathematical Tables 426

IV. Important Formulas and Symbols. ... 443

V. List of some important Publications containing Statistics. 446

VI. Some important Books on Statistical Methods. 451

ANSWERS 453

INDEX 455

CHAPTER I

INTRODUCTORY

The word 'Statistics', derived from the Latin word *status* meaning a state, is found used in singular as well as in plural. Its meaning, however, changes with the change in its number. Statistics (*singular*) refers to the subject of statistics as a whole and is as such a branch of Applied Mathematics. When used in plural it signifies numerical data or statement of facts systematically arranged, From *status* appear to have originated the Italian word *statistics* meaning a person well-versed in state matters but occupying a subordinate position and who was expected to be of help to high government officials responsible for formulating the policies of the state. In this sense, this word is found used in the seventeenth century in the writings of Shakespeare* and Milton. In the 'Elements of Universal Erudition' by Baron J. F. Von Bielfeld translated by W. Hoofer H. D. (London 1770) there is a chapter on 'Statistics' wherein it has been defined as the 'Science that teaches us what is the political arrangement of all the modern states of the known world.' In 1787 Prof. E. A. W. Zimmerman of Brunswick referred to statistics in the preface of his book, 'A Political Survey of the Present State of Europe' as that branch of political knowledge, which has for its object the actual and relative power of the several modern states, the power arising from their natural advantages, the industry and civilization of their inhabitants, and wisdom of their governments. The phrase, 'Statistical enquiries' appears first in the Circular issued in 1790 by Sir John Sinclair, Editor of the Statistical Account of Scotland, meaning 'enquiries respecting the population, the political circumstances, the productions of a country and other matters of the state.' The term 'Statistics' is said to have been first applied to this line of study by Gottfried Achenwall, often called the 'Father of Statistics'; while the name of 'Political Arithmetic' was given by William Petty. The meaning of the word statistics has thus undergone changes from time to time.

---

* I do believe,
Statists though I am none, nor like to be
Cymbelline Act II Sc. IV.

† Their orators though then extolest; as those
The top of eloquence, *statists indeed.*
And lovers of their country as may seem,
Paradise Regained Book IV.

Meaning and Scope of Statistics:—The word statistics conveys two different meanings; in the singular it refers to the subject as a whole and in the plural, to numerical data. Webster defines statistics as, "Classified facts respecting the condition of people in the state—especially those facts which can be stated in a table or tables of numbers or in any tabular or classified arrangement." This definition fails to take into account all the aspects of statistical method and limits it to the condition of people in the state. Statistics now covers a far wider field as it is applied to the study of data concerning problems relating to astronomy, meteorology, biology, etc. Dr. Bowley's definition, 'Statistics is the science of the measurement of the social organism, regarded as a whole in all its manifestations' is also defective as according to this, the scope of the science will be restricted to man and his activities. As said already, the science has a wider applicability. The other definition of the same author, 'Statistics is the science of counting' is also not correct. Statistics is concerned not only with counting but with tabulating, summarizing, drawing graphs and similar other things. Again, the word counting conveys an idea of exactness but in statistical operations estimates play an important part particularly when the number of observations is large. When we think of agricultural crop statistics, we cannot vouch for their exactness as they are more of estimates. The definition, 'Statistics may be rightly called the science of Averages' is also not appropriate as it takes into account only one step in the process of statistical operations. In this respect the definition given by Lovitt and Holtzclaw, 'Statistics is the science which deals with the collection, classification and tabulation of numerical facts as a basis for explanation, description and comparison of phenomena' is far more comprehensive but with a view to mention the scope in detail it can be improved by defining statistics as 'the science and art of handling aggregates of facts observing, enumerating, recording, classifying and otherwise systematically treating them so that master facts or principles or laws lying behind or contained in the aggregate are made comprehensible to the mind and become along with the results of other methods of investigation, data for reasoning, the drawing of conclusions, the making of decisions and the determination of policy'.

The word Statistics in plural refers to numerical data and so in this case, 'Statistics are numerical statements of facts in any department of enquiry, placed in relation to each other'. It must, however, be noted that isolated figures having no connection with each other cannot be styled as statistics as they are not the result of any enquiry and also do not bear any relation to each other. Simply giving numbers like 27, 47, 58, 85, 12, 20 does not constitute statistics. But if we collect figures of prices of wheat for a week and those of the quality sold at those prices we get statistics of prices as well as of the sale of wheat.

Characteristics of Statistics:–Statistics (meaning numerical data) possess important characteristics which have been included in the definition framed by Secrist, 'Statistics', according to him, 'are aggregates of facts affected to a marked extent by multiplicity of causes, numerically expressed, enumerated or estimated according to reasonable standard of accuracy, collected in a systematic manner for a predetermined purpose and placed in relation to each other'. Statistics are thus aggregates of facts and so a single figure of sale or of exports cannot constitute statistics. They are affected to a marked extent by multiplicity of causes. Statistical phenomenon can rarely be the subject of change due to a single cause only. The reduction in yield of wheat may be due to want of adequate rainfall, want of proper care on the part of agriculturist or want of adequate manuring. Statistics are required to be expressed in numerical quantities. There cannot be statistics of qualitative expressions. We have to assign marks to the answers of examinees and it would serve no statistical purpose if we mark them as good, bad fair, tolerable, etc. The enumeration or estimation done must be according to reasonable standards of accuracy. What is a reasonable standard will depend upon circumstances. While recording imports of gold, ounces will be taken but not so in the case of steel. Again, the standard of accuracy desired will depend upon the nature of results expected. Statistics must be collected in a systematic manner. There must be proper planning as well as organization to avoid duplication and confusion. Statistical operations cannot be carried out satisfactorily unless their purpose or object is determined. If we have to construct an index number of prices we must decide its purpose at the outset. They must also be capable of being placed in relation to each other. In short, they must possess the characteristic of comparability. Stray numerical facts are not comparable unless they are homogeneous. For achieving satisfactory results statistics should possess these characteristics in as high a degree as possible.

Importance and Value of Statistics:—Though a statistician is sometimes talked of as a *figuring fool* because of his spending a good deal of time in long and abstruse calculations of doubtful value the utility of statistics is greatly appreciated by economists, scientists, social workers, commercial magnates and others. As early as 1890 Dr. Alfred Marshall while giving evidence before the Committee of Census said, 'Statistics are the straw out of which, I like every other economist, have to make bricks'. For a scientist the importance of statistics is very great. In taking observations and ascertaining them this science does play a very important part. For a business executive statistics are an aid in supervising the work of his employees and in planning his sales and purchases as well as production. Because of this growing importance a statistician these days occupies in business an important position between a business executive and a cost accountant. During the last war the government was greatly helped by statisticians in planning several branches of administration especially in the war department. Even in the days of food scarcity, the efficiency of the food department depends to a great extent on statistics.

The Value of Statistics may be broadly summarised as under: (1) Statistics helps in presenting an unwieldy mass of matter in a simple, definite and classified form. When statistics are collected they are in the form of an unwieldy mass and as the human mind cannot ordinarily grasp the matter in such a form it requires to be treated statistically. If wages sheets of all the two thousand workers in a mill are given, it is impossible to form a correct idea about them from these sheets. But if they are tabulated and summarised, definite information will be easily available. It has, therefore, been said, ' By the use of statistics we obtain from milliards of facts the grand average of the world'. It was with the help of statistics that the British Government during the last war could obtain from the census papers definite information about its man-power for enforcing conscription according to certain age limits.

(2) Statistics enlarges individual mind :—Ordinarily an average individual does not go deep into the details of various matters by scrutinising them but if they are presented in a suitable and handy manner they can be grasped and remembered by persons even though not particularly interested in them. If birth and death rates, exports and imports, per capita income and such other things are properly presented they will attract the attention of a large number of persons who will not only take interest in these things but will even remember them. It is because of this that Whipple has aptly said ' Statistics enable one to enlarge his horizon. '

(3) Statistics furnishes methods of comparison :—Statistics provides us with various methods of comparison such as different forms of averages ratio, graphs coefficients and diagrams and thereby it becomes very easy to compare data and to weigh and judge them in the right perspective. Income per capita in different countries, ages at deaths, incidence of taxation, literacy and unemployment can be compared with advantage. Not only comparisons can be made but by such process improvements can be effected in the economic social, commercial as well as political condition of different parts of the globe.

(4) Statistics helps in ascertaining conditions of relationship :—There are some factors such as rainfall and yield, supply and prices which bear relationship in their movements. With the help of statistics, effort can be made to establish correlations

between such factors. Once the normal degree of correlation is ascertained the effects of changes in one can be easily seen in the other. This factor affords us a great help in solving many problems of life—economic, scientific as well as medical.

(5) Statistics, considered as the logic of figures, assists us in arriving at correct views based on facts. By studying statistics of the past we come to know what has happened in the past and under what circumstances, and by analysing these cases and by tracing the causes leading to them we are enabled to forecast the future with a fair amount of accuracy. It is no doubt true that statistics only indicates trends and tendencies but even then this is of great assistance to the economic as well as commercial world.

(6) Statistics endeavours to give material for the business magnates as well as to administrators to serve as a guide in planning as well as in shaping policies for the future. When the budget is in deficit the finance minister finds it handy to arrive at the scheme of his taxation by making use of statistics. Similarly for fixing sales quotas for the various salesmen working under a business executive he finds statistics as a valuable guide.

(7) Statistics is of great help to other sciences such as biology, agriculture, medicine and meteorology as by adoption of suitable statistical methods determinable facts can be easily weighed in the balance to assist in judging possibilities from ascertained results and tendencies. The whole doctrine of heredity and evolution is largely based on statistics. Mortality tables form the basis of life assurance and are mainly due to study of vital statistics. Without the aid of statistics it would be difficult to make a rapid progress in meteorology. The various experiments carried out in the science of agriculture for improving the yield as well as variety of different cereals owe their success to a great extent to statistics. For studies in economics too it is of great value. According to F. C. Mill, ' Mathematics and its offspring, statistics and accounting are the powerful instruments which the modern economist has at his disposal and of which business, through the development of research agencies and methods, is making constantly greater use.'

Statistics and Commerce :—The part played by statistics in commerce and industry is equally great. With the increase in business and productive activity it is getting increasingly difficult for a business executive to control and supervise the working. With the help of statistical table and more so with diagrams and charts inner controlling master facts can be easily ascertained as well as borne in mind. Statistics are, therefore, said to be the eyes of administration. Bearing on the past statistics it has become easy to lay down standards by which the current performances can be valued and compared. This has also given facilities for fixing performances for the future based on the past and thereby the work can be easily supervised. By studying statistics and effecting comparisons it has become easy to lay down the likely tendencies which guide businessmen. This will not only save them from would-be losses but bring in profits due to proper forecast. Mastery over statistical methods gives the necessary insight to do market research and to formulate a business policy. Purchases can be made in favourable times to meet the volume of estimated sales based on the statistics as well as on market research. Seasonal changes in demand and supply can be estimated with a fair amount of accuracy and steps can then be taken to adjust the stock to the circumstances. Similarly boom or slump can be forecast and the business activity regulated accordingly with a view to guarding against any unfavourable situation. It may not be an exaggeration to say that statistics helps us in determining laws governing industrial operations.

Limitations of Statistics :—Statistics is a very useful tool for a scientific enquiry as the study of statistical data necessarily involves the scientific ascertainment of facts or the relation of facts from a mass of material. But as stated by Newsholme, ' it must be regarded as an instrument of research of great value, but having severe limitations which are not possible to overcome and as such they need our careful attention'.

The subject of enquiry may or may not be capable of numerical measurement. Rainfall, output of yarn or heights of persons can be measured quantitatively but this is not possible in the case of problems relating to intelligence, poverty, health, friendship or culture. Intelligence or poverty cannot be measured quantitatively. Intelligence of two branches of students can be compared indirectly by taking results of their periodical tests. Efforts can be made to have an idea of poverty by finding out persons taking relief or assistance. In the study of the phenomenon of such types the statistical aspect is bound to be the subsidiary to other considerations. Statistics mainly deals with the averages and so in them individual peculiarities are submerged. No. particular difficulty will be experienced in studying only the trends but if individual peculiarities are of material importance the results obtained cannot be satisfactory. The average consumption of cloth per head ascertained by dividing the total consumption by the population will not give a real idea of the shortage of cloth experienced by the poor. The average is raised by the extravagant use of cloth by the rich. Due to the difficulty of studying all the cases statistical results are ascertained by samples. If the selection of samples is free from bias the result will be near the average due to the operation of compensating errors. If there has been deliberate sampling, the errors will, on the contrary, accumulate and statistics cannot, in that case. be reliable. They are more of estimates than exact statements. This is more so as is in statistical studies, particularly business and economic, the scope for carrying experiments under controlled conditions is limited. Unlike a chemist a statistician cannot work in laboratory where at a time only one factor can be studied by controlling other disturbing elements. Statistical results are also likely to be used without their context and consequently the picture presented may fail to represent the true aspect. No doubt the science of statistics presents facts but, it is limited to the numerical aspect only and so tendencies deduced need consideration in conjunction with other relevant factors regarding the economic, social or political aspects. Lastly the, danger of misuse of statistics is very likely when they are to be handled by persons who do not possess expert knowledge of statistical methods or by interested persons.

Distrust of Statistics :—The science of statistics is ve useful but the value and uses of statistics lie not in the figu themselves, but in the deductions to be obtained from them. D to the ignorance of the public as well as the misuse of statis

by interested parties to serve their ends this useful science has been blamed. According to Mark Twain, 'There are three kinds of lies–namely *lies, damned lies and statistics*–wicked in the order of their naming'. It is often said, 'With statistics anything can be proved'. Karston in his book on 'Graphs and Charts' writes 'The statistician is sometimes looked upon as one whose acquaintance with figures is so very intimate that he can readily take liberties with them abuse them present them in a false light and deceive the layman. In this view he is a little more than a common trickster performing leger-demain with numbers his magical results to be idly wondered at but not to be trusted'. There is, however a section of people which strongly believes in the old proverb 'Figures won't lie'.

This distrust of statistics cannot be attributed to this science. This is mainly due to the credulous nature of the people and the deliberate manipulation of statistics by unscrupulous persons to support their views unwarranted by facts. Statistics cannot bear on their face the stamp of their quality and without a very careful scrutiny it is not possible to differentiate between a table compiled by an industrious and conscientious statistician by spending hours over the various statistical operations and one based on unreliable data and that too collected carelessly. Even if there be some glaring inaccuracies a casual observer is not likely to trace them and rarely cares to enquire into the competence of the author.

This manipulation of facts is often done by quoting statistics compiled for a certain purpose to prove a very different thing or by quoting figures without proper context. The results obtained by a statistical survey of a small group are also used to show the general trend. As already explained there are limitations of statistics but they are not purposely taken into account and statements are made without giving reference to them. It is also not uncommon to find persons applying only the favourable results and thus making unwarranted statements in the name of statistics. This distrust is also due to fallacious comparisons and statements loosely made by interested parties. Really speaking the science of statistics and its technique if used properly and without any sinister motive, are very useful and can on no account be blamed as, 'Figures don't lie, but liars figure.' If a sword is used for killing an innocent person instead of protecting him it is not the sword but the person who is to be blamed for its misuse. This distrust is, however, slowly decreasing because of the increasing interest evinced in the study of statistics and the recognition of its limitations as well as improvements in statistical methods.

Main Divisions of Statistics :—The domain of statistical studies can be conveniently classified in two main divisions or branches as under :—

1. **Statistical Methods** :—It can be considered as a branch of mathematics as it attempts to formulate definite rules of procedure which can be used in handling groups of data of different varieties. It is by the process of statistical method that large masses of figures are dealt with particularly for elucidating quantitative data affected by multiplicity of causes. The scope of statistical method includes mainly collection, tabulation and analysis of data by various methods as well as their interpretation. In short, it is a distinct procedure in dealing with facts and their relations. The important functions of the statistical method are, giving in a condensed form, the description of the whole series of observations and examining the various hypotheses made about the observational data.

2. **Applied Statistics** :—It is an application of rules and formulas laid down by the methodologists to concrete facts as they exist and its important branches are (1) Biometry dealing wish measurment of living beings, (2) Psychometry dealing with mental phenomenon, (3) Vital statistics and Demography concerning population, (4) Administrative statistics dealing in government matters, (5) Social statistics concerning social phenomena and (6) Economic statistics including those relating to business. Applied statistics exemplify the application of the principles of specialized facts and problems.

Applied statistics can be further classified as *Descriptive* and *Scientific*. The former deals with known data or records relating to the past or present and as such Business statistics forms part of it. The latter is mainly used for determining physical and psychological laws by investigating the data collected for descriptive purposes by statistical methods. Scientific statistics is the ultimate goal as it puts to account the rules of statistical methods as well as the data collected for descriptive purposes.

Role of Statistician:—The scope of the science of statistics as shown already is sufficiently wide. It includes the development of mathematical theory underlying statistical methods, statistical methodology and its application to different fields. It is hardly possible for anybody to be thorough in all the branches. Besides, he is faced with several difficulties. He possesses no laboratory, has to deal with data regarding human objects having individual bias and has to depend upon samples which may or may not be truly representative. He is concerned with the collection, analysis and interpretation of statistical data and has to work on the principle of division of labour to achieve speedy as well as efficient results. Under the circumstances, he is not expected to work miracles. His work will depend upon the material and co-operation he can command. To succeed well in his mission he must be an impartial and conscientious worker without personal prejudices but with sound common sense. He should particularly note that 'Excrcise of judgement based on wide familiarity with the situation and greater caution in the statement of conclusions from the data, are the beginnings of wisdom in statistical work.'

Exercises.

1. 'Statistics is the science of counting'. Criticise this definition giving arguments in support of your answer.
2. Mention the different branches of statistics and describe each of them.
3. Establish the importance of the science of statistics giving its uses particularly to commerce.
4. Reconcile the two statements give below :—
   (i) 'With statistics anything can be proved'.
   (ii) 'Figures do not lie.'
5. Trace the gradual development of the science of statistics.

6. 'Statistics are the straw out of which I like every other economist have to make bricks'. Discuss this statement of Marshall.

7. Define statistics and state the difficulties that a statistician has to encounter in his work.

8. Discuss the scope, utility and limitations of statistics.

9. 'Statistical methods are most dangerous tools in the hands of an inexpert'. Support this view by giving suitable examples.

10. Give the important uses and limitations of statistics and show its relation to economics as well as to mathematics.

11. 'There are three degrees of comparison in lying. There are lies, there are damned lies and there are statistics.' Discuss this statement and state with reasons how for do you agree with it.

12. 'Sciences without statistics bear no fruit, statistics without sciences have no root'. Comment.

13. How far is the common saying, 'Statistics can have anything correct ? Discuss.

(1936–Oct).

CHAPTER II

STATISTICAL INVESTIGATION

Though it is comparatively of late that the science of statistics in its present form has attained such a great importance in all spheres of life, political, social, economic, industrial as well as scientific, statistical compilations can be traced even to the ancient times. It was as early as 3050 B. C. that statistical investigations were undertaken in Egypt regarding its population and wealth in connection with the construction of the famous pyramids. 'Indika' and 'Anabasis', two works of Arrian who came to India with Alexander the Great as well as the writings of Megasthenes speak of the minute statistical collection at that time in India. In China too a detailed statistical description of the provinces was compiled in 1200 B. C. But it is in recent times and more particularly since the First World War that various problems of diverse nature are coming to the forefront because of the increasing complexities of the modern world. Marketing, production, consumption, banking, finance, transport and other problems have attained far greater importance because of their international as well as interdependent character and research has become a matter of necessity in these as well as in other matters. Physicians want to investigate the efficacy of a certain medicine; chemists desire to statistically test their synthetic compounds. The success or otherwise of inoculation is seen by statistical methods. Manufacturers rely on statistical research to serve as guide for production and marketing. Ministers depend upon statistical material for budgetary and other schemes and so do bankers with a view to regulating their policy of fixing the amount of their cash reserves as well as rates of discount and interest. In short, the use of statistical investigation and research is getting popular day by day.

Stages in a statistical investigation:–Whenever any problem is taken for statistical investigation and research, it involves a number of operations such as defining the problem, organizing and planning an enquiry, determining the degree of accuracy to be aimed at, fixing the units, collecting and tabulating the data presenting the same in a summarised form and preparing the report after proper analysis and interpretation. All these operations which are worked with different kinds of statistical methods can be classed into stages of steps: (1) The Enumeration or Collection of data (2) The Tabulation, (3) The Summarization of the data by calculating averages, constructing graphs or pictograms and (4) Critical examination based on the statistical investigations and research. All these stages have been described at length in succeeding chapters.

Preliminaries to the collection of data:–Before we actually enter into the first stage of statistical investigation, i. e. collection of data it is necessary that every feature of the problem or enquiry should be carefully examined so as to obtain the maximum benefit from the investigation to be undertaken. Failing this, in the midst of the investigation it is likely to be noticed that important items have been overlooked and unnecessary details on certain points are being collected. The steps to be taken in this connection include matters such as defining the problem, planning as well as outlining the purpose and scope of the enquiry, laying down the standard of accuracy, the mode of treatment, defining statistical units and determining the available sources of information.

Defining the problem:—To start with a clear and concise statement of the information to be collected regarding the problem must be made, as slight deviation in its scope will mean at least a partial change in the enquiry. If the problem of unemployment is taken for investigation it is better that it is made definite by limiting it to a particular set of people such as educated or uneducated, skilled or unskilled labour. If the problem of wages is being investigated it is necessary to make it definite by laying down whether money wages or real wages are to be considered and if other advantages like overtime earnings, free education, etc. are to be taken into account or not. In case of the problem of profits of companies it is equally necessary to decide whether gross or net profits are to be considered. The advantage of this will be that irrelevant matter can be easily neglected giving thereby more importance to necessary data and guarding against the omission of matters of vital importance.

Planning:—Planning is the most essential requisite in a statistical investigation in the absence of which time as well as energy and money are likely to be wasted. Before actually starting investigation the available existing sources of information should be thought of; otherwise it will be noticed that unnecessary energy has been spent on collecting the material that is already in existence In some cases the requisite material may be in existence in a different form and needed only shapping. In many instances it is better to carry on the investigation in such a manner that more important questions regarding the problem can be taken up first as it is likely that it may be seen from the answers received in some cases that carrying further the investigation would result in no advantage. Thus, a considerable amount of energy, time and money can be saved by dropping the project in hand or by making necessary changes. At the time of planning all possible sources of error will have to be anticipated and guarded against as a proper precaution taken at the time of planning would simplify the work in the future. For instance, when a census of population is planned the day chosen should be one on which most of the persons are likely to be in their places. In India particularly in marriage season or on days of festivals or fairs many persons go from one place to another and thereby the work of census, if fixed on such days, is likely to be more troublesome as well as liable to inaccuracies. A statistician has no laboratory like a chemist or a physicist and so he has to be more careful in matters of planning and the details connected with it, to avoid difficulties in future operations.

Purpose and scope:–The most essential thing in a statistical investigation is the knowledge of the purpose and scope of the enquiry as on this depend the various factors such as methods of enquiry, degree of accuracy and the extent of the data to be collected. It will serve as a guide in solving difficulties arising in the process of compilation. A statistical enquiry may be undertaken with a view to obtaining description or scientific data. In the former case the material is dealt with primarily from the historical point of view while in the other the aim is to provide material to verify some hypothesis or to establish a law. The purpose again can be general or specific. It may be the policy of a government or a business office to maintain statistical records on a comprehensive basis for general purposes. If it is for a specific purpose the data to be maintained and collected should be based on the specific object in view. The decennial census, statistics of trade as well as production can be cited as examples of statistics maintained for general purposes. In the compilation of this type of data the danger to be guarded against is the likelihood of cost becoming disproportionate. In many cases it may be found on careful consideration that a sample enquiry is likely to serve the purpose without much loss of efficiency. The question will, however, be discussed at length in a separate chapter.

Degree of accuracy:–Absolute accuracy even though it can be attained is not necessary to aim at as it will be very often at a very high cost and without considerable addition to the required standard of accuracy. Again the desirability of the degree of accuracy will depend mainly upon the purpose of the enquiry. When we are investigating the sources of imports it would be sufficient if figures in thousands of rupees are taken. But if the research pertains to the rates of interest prevailing in the market even two digits of decimals could be taken with advantage. Very often it is better to prepare an approximately accurate report quickly than to try to compile it by taking a long time with a view to aiming at a slightly higher accuracy. It is equally important to think of the inaccuracies that are likely to creep in because of clerical and other types of errors.

The unit of statistical data:–A statistical unit which is at the basis of the compilation should satisfy the following requiremeuts :—

(*a*) Its definition must be simple, concise and clear. In quantitative science a precise and unambiguous terminology must be adhered to. Change of definition or different enumerators interpreting it in different ways will vitiate the results. At the outset the task of defining units appears to be simple and easy but actually it is otherwise. For instance, when persons are to be classed as married or unmarried, in which class should a widower or a widow be placed? When the price is to be noted, which price, cost, selling, retail or whole-sale should be taken? A great difficulty is experienced in interpreting words like profits, literate, house in the absence of clear instructions as to the method of treatment in the compliation of the data particularly as different purpose often necessitate change in the scope of the definition. In short, the unit must be specific and unmistakable.

(*b*) The unit taken should be stable and if it is not so some convenient stable standard should be adopted. Suppose we want to use a rupee or a dollar we will have to fix as standard unit its value on a particular date and day to day values can be converted on that basis. In India weights vary from place to place. A seer is of 40 tolas in Gujarat while in some parts of U. P. it is of 100 tolas. To avoid confusion some standard unit will have to be fixed for the guidance of enumerators.

(c) The unit must be appropriate to the enquiry undertaken and also capable of correct ascertainment. While enquiring about the labour employed the weaving department of a mill will think of the labour of that department but from the point view of a mill the labour will include the number of persons working in the various departments as well as in the office. The unit will, therefore, depend upon the nature of the enquiry.

(*d*) There must be homogeneity as well as uniformity in the characteristics of units as, in the absence of this important quality, comparison will be very difficult. Particularly when average is to be taken, homogeneity is most essential. If this is not the case, the difficulty will have to be got over by subdividing the data into groups or expressing dissimilar data in terms of equivalents of the selected unit. For instance, the data for compilation of an average of the prices of tea sold in the market, where three grades of tea are sold at different prices, is heterogeneous. In such a case we will have to subdivide the data into three groups or classes; price for grade A, price for grade B and price for grade C or to express the dissimilar units in terms equivalents of the selected units i. e. by fixing some proportion of these three grades of tea. This quality of homogeneity is particularly important in studying price changes, growth of business, bank clearings extending over a period of time as changes in conditions make it difficult to make proper comparisons.

(*e*) The units taken should have representative character. In case all the items are not likely or convenient to be studied the units selected must be of representative type as, if otherwise, the result obtained will not give a true picture of the group.

Types of statistical units :—These units can be classified as (1) units of enumeration and (2) units of analysis and interpretation. The former are concerned with the measurements made and are of two types: simple and composite. *A simple unit* implies a collection of characteristics that usually occur or are thought of together. Its meaning is general and it simply distinguishes classes; a yard, a rupee, a house denote simple units and are simple to define. *A composite unit* is formed by adding a qualifying word or phrase to a simple unit and then its scope is to some extent limited e. g. a ton-mile or a passenger-mile. The amount of a ton-mile can be arrived at by multiplying the number of tons by the number of miles over which they have been carried.

Units of analysis and interpretation are those in terms of which comparison of the data is made and include ratio, coefficients and rates. *A Ratio* is a way of expressing the percentage difference upon some quantity taken as standard. If the production of a field is 250 as compared to 100 units of the previous year it can be expressed in the ratio of 100 : 250 or 10 : 25 or 2 : 5

A coefficient takes the form of a comparative statement and the comparison may relate to space, to time or to conditions in either of them. The production of jute may be shown in maunds relating to a province, to a district or even to an acre. Wages may be shown in relation to time by expressing them as so many rupees for days or months. They can also be expressed as so many rupees per week or per day. We also show income of a particular province or country in money units but relating to the whole of the population and also speak of income per head. When there is a comparison of quantities of some kind we use the term ratio. If a village has 250 literate and 1000 illiterate inhabitants the ratio of literate to illiterate will be shown. But when the two i. e. the numerator and the denominator are of different kinds it is designated by rates such the death rate calculated by comparing the number of deaths to the population. As it will be seen in this case the comparison relates to quantities of different kinds. Incidentally it should be noted that the death and birth rates are expressed not as percent but as per mille i. e. per one thousand.

In the end, it is very important to note that for a successful statistical investigation the statistical data to be collected must possess the essential characteristics of comparability, homogeneity, completeness, stability, representativeness and freedom from personal bias. What comparability can there be between the profits of two companies under investigation if in case of one we take profits of 1945 and in the other those of 1935 ? We shall have to take for this purpose either profits of 1935 or 1945 in each case. Homogeneity is another characteristic of importance. For this purpose all the statistical facts have to be analysed into groups containing only those units that are similar. For ascertaining the average wages of labour in two mills there cannot be homogeneity if in one case we take the average of wages of only male labourers while in the other we take into account wages of both male and female labourers. Incomplete statistical data will vitiate the results and the same will be the result if data without stability or representative character is collected. To ensure stability either a convenient stable standard has to be adopted or an average should be taken. To bring in representativeness the easy way is to guard against isolated cases and to take care in sampling the data. Personal bias is a dangerous factor in the collection of statistical data and this as well as sampling including errors have been fully discussed in a separate chapter to guard against mistakes due to bias and sampling in the collection of data.

Exercise

1. What is a statistcal investigation ? Discuss its utility to various types of people.

2. Describe the main stages in a statistical investigation.
3. Mention the points that should be carefully considered before starting a statistical investigation.
4. Emphasise with illustrations the importance of an exact definition in statistics.
5. Planning is one of the essential requisites of a good statistical investigation. Justify this statement.
6. Discuss fully the essential requisites of a good statistical unit.
7. Write short notes on :—
   A composite unit, degree of accuracy, units of analysis, personal bias.
8. If you were to conduct an enquiry into the payment of wages in an industry on what points would it be necessary for you to be clear before actually commencing the investigation work.

CHAPTER III

SAMPLING AND APPROXIMATION

In the study of statistics it is necessary to distinguish between statistical description and statistical induction. With the help of various statistical devices such as averages, dispersion skewness we attempt to describe with fair amount of accuracy a large mass of quantitative data by replacing the number of individual cases by a limited number of measures which bring out the characteristics of groups as a whole. When the results obtained are confined to the cases actually studied these statistical measures describe certain features or relationship of the data and the description can be relied upon to a great extent as an accurate description of the given characteristics. The real problem arises only when we mean to generalize the conclusions arrived at by the study of some cases and attempt to apply them to those not included in our original study. This process of statistical induction or inference, i. e. of generalizing the results obtained by studying sample cases and assuming them to be applicale to the whole data leads us to the discussion of sampling and its reliability.

Census and Sample :—When a statistical inquiry is to be undertaken the immediate problem is whether the investigation is to be concerned with the whole group, commonly designated in statistical parlance, as '*Universe*' or '*Population*', or with only a part of it. These two methods of an enquiry are distinguished by the words *Census and Sample.* In census we collect data regarding the universe or population and so the problem of drawing inferences about the whole group by taking observations of the part does not arise. In the case of a sample enquiry we try to generalize in terms of the whole group though the facts assembled relate only to a part of it. No doubt the results obtained by a census are likely to be more reliable than those by taking samples. Yet it is, not normally expedient or possible to obtain information of the whole population, or 'universe.' In case we want to study the economic condition of the inhabitants of a village it will entail a very heavy drain on time, energy and money if we were to study each case but the results obtained by this method will not be materially different if we were to achieve them by the sample method. If we were to select persons at random we shall find that the data will be of a representative character, provided a sufficiently large number has been taken. This holds good because of the scientific principle, '*Organised' knowledge is representative in character'.* It is a matter of common knowledge that lakhs of rupees are spent on the decennial census besides taking honorary services of several thousands. Such an enquiry is not possible every now and then and so even the government takes the census only once in ten years. Besides, as a large number of persons is required for an enquiry it is neither possible to obtain a sufficient number of trained personnel nor to train the enumerators in a short period with a view to minimise the likely errors. In short, considering the question of time, energy, expenditure and skilled as well as efficient labour the sample method is found to be comparatively more advantageous than the census only if sampling is based on the principles of *Statistical Regularity* and of *Large Numbers.* These two principles are usually referred to as laws but strictly speaking they are tendencies that operate universally.

Law of Statistical Regularity :—This principle. which is derived from the mathematical theory of probabilities, enunciates that if a sufficiently large number of objects be selected *at random* from a large group they tend to possess the characteristics of the larger group. It must be carefully borne in mind that for the operation of this important principle the selection of samples must be made at random and from objects every one of which having an equal chance of being selected. This can be done by taking samples by a lottery, by drawing them haphazard or by arranging the objects and choosing certain numbers like 12, 24, 36, 48 and so on. The point aimed at is that there should not be deliberate or conscious selection as it is less dependent upon the theory of probabilities. Again, the number of samples taken must be large enough as with the increase in number the chances of samples conforming to the group increase and then the information collected will be more dependable as the reliability of the indication of the sampling is proportional to the square root of the number of items included. Because of this it has become possible to present a sufficiently accurate picture of a group without following the costly and labourious method of census i. e. of having a complete enumeration of all the objects. As a matter of caution it should be noted that the results obtained from the sample data will not be *exactly* equal to the results got by investigating the whole population but they will be very nearer to them depending upon mainly the large number of samples and also the method of sampling.

Law of Inertia of Large Numbers:–This is a corollary of the Law of Statistical Regularity and enunciates that if the numbers involved are large the total change is likely to be quite insignificant. In other words the larger the groups or aggregates of data the higher will be the degree of stability. This is due to the fact that in a larger group when some objects will vary in one direction others may vary in the opposite direction a few remaining without registering any change. All this means that changes in either directions would to *a great extent* compensate one another and such changes are more in a larger group than in a smaller one unlsss by a rare chance most of the movements are one-sided. By this it is not meant that there would be no changes or movements over a period but it only shows that such changes, when a large group is concerned, will be more gradual as well as regular.

Attributes of s good sample:—With a view to obtaining satisfactory results by taking only samples and by not taking a complete census of the universe it is incumbent that the samples chosen should possess the essential attributes. Every object in the universe must be independent and must have an equal chance of being taken as a sample. If otherwise, samples are likely to be drawn more from one group which will defeat our purpose. The samples should be such as would really represent the mass forming the whole population and not only a part of it. To guard against this danger the easier course is to group the data properly as this will not allow lifting unrepresentative samples.

Personal bias very often comes in the way of right type of sampling and this can be got over by random sampling or by some other method. Unconsciously most easily accessible objects are selected to avoid extra work but therein lies the risk of only a particular type forming a majority in the number of samples taken. This is far more true in the case of returns or samples voluntarily tendered as they are very often from interested parties. Lastly, as already emphasised, the number of samples taken should be as large as possible because with increase in the number of samples there will be an increase in the accuracy in the results to be achieved. 19.3.57

Methods of sampling:–The methods commonly used for selecting samples from the 'Universe' or 'Population' are (1) Extensive, (2) Representative or Deliberate and (3) Random or Chance sampling,

(*a*) **Extensive sampling**:—In extensive sampling effort is made to obtain information about as many objects as possible and so this process leads to almost complete enumeration. As all cases are not normally covered we call it sampling in the strict sense of the term though it is more an enumeration of the group than sampling, As is natural in this method all the cases about which information can be *easily* obtained are included with the result that it does not become really representative in character in the absence of the inclusion of the information about a section of cases for which more effort is needed, This may or may not be due to any bias but all the same this defect is very likely to exist in the result obtained.

(*b*) **Representative sampling**:—In this method effort is made to get over the want of representative character by conscious, purposive or deliberate selection. The person in charge of the investigation selects cases which appear to him typical or representative in character. Though by this method we try to get over the defect of the previous one it is very likely that the investigator may consciously or unconsciously select such cases as may produce results desired by him. But if he is not allowed to exercise his judgment and samples are selected at random most probably the results may be more reliable. This is, therefore, a desired method if it is being worked out by persons who are seekers after truth but, if otherwise, it will be a dangerous sharp-edged tool of mathematical interpretation in the hands of an average investigator.

(*c*) **Random or chance selection** :—In this method there remains no scope for bias of any type as the samples are selected by a draw in a lottery, by taking samples at some stated ranks in the population say, every eighth or at some other interval or by picking them here and there without any predetermined idea. In case of a lottery every item is given a consecutive number and slips are made out for each number and they are then freely mixed to change the order completely. Slips are then picked up and those cases shown by the numbers chosen are investigated. The other method is of arranging the data in an ascending or descending order or in groups and then choosing numbers at stated intervals say 7th, 17th, 27th, 37th and so on. In *stratified sampling* the objects are divided into convenient groups having as far as similar characteristics and then samples are selected by taking equal proportion of items from each group instead of random sampling from the whole universe. This is likely to yield better results particularly in social and economic surveys. In case of haphazard picking of samples there is no particular method except depending fully upon the theory of probability. However, it is necessary to see that only those cases quite close at hand are not taken to the exclusion of others as this will defeat the very object of random sampling. In random sampling particular care must be taken to see that at the time of sampling all objects possess equal chance of being selected and that a sufficiently large number of objects is taken for investigation to allow a fairplay to the two laws of Statistical Regularity and Inertia of Large Numbers.

Reliability of Sample:—The reliability of the sample taken can be seen by the "*Stability test*'. In this the samples taken are divided into two equal parts by random sampling and then it is seen whether or not the characteristics of the original samples are depicted to a reasonable degree in the divided parts. The other way is of selecting from the original data a new and unbiassed sample equal in size to the sample already taken for comparing the characteristics.

APPROXIMATION

Statistics is mainly based on estimates and so the accuracy of accounting or mathematics cannot be expected. In the first place it is very rarely possible to collect information regarding every object in the ‘universe’ or ‘population’. When we are preparing crop statistics or statistics of yield it is almost impossible to measure every inch of ground sown under a particular crop nor is it easy or probable that every seer of yield will be accurately measured and recorded. Even with the best intentions this is not practicable and an attempt at this will mean unnecessary waste of energy and money without any adequate compensation. This is achieved mainly by estimates. Even if we think of the decennial census accuracy remains only the cherished object. When millions of persons are to be counted and the collection of information in some cases is left out inaccuracy does creep in. What is to be guarded is a reasonable degree of accuracy by resorting to more up-to-date methods and aiming at as high a degree of accuracy as possible which will mainly depend upon the circumstances of each case. For instance, in the case of the production of gold ounces will be cared for while in the case of the production of iron even a pound will not be given much importance.

In statistical work approximation plays an important part along with estimation. Many investigations are based on estimates and their result worked out on approximation may very often be of great importance from the point of view of ease as well as clarity. If the figures of export of cotton and its consumption in the country are to be collected it serves no purpose to put them on record to the last digit of ten or of hundred. Approximate figures correct to the nearest hundred or even thousand maunds in some cases will be appreciated as it will reduce the calculation work without much loss in accuracy in the results. The process of approximation consists in showing the data having large figures more clearly without much loss of efficiency and thereby faciliting further calculations. The output of the colliery A can be shown as 21,75,000 when it is actually 21,75,123 tons and 12,62,000 of the colliery B when it is recorded as 12,62,112 tons.

The degree of accuracy can be shown by adopting one of the following ways :—

(*a*) The quantity of wheat sold is 5570 maunds in round number.

(*b*) The quantity of wheat sold is 5570 maunds plus or minus an amount not exceeding one–half. Mathematically it can be shown as the quantity of wheat sold is 5570 ± ·5 maunds.

(*c*) The quantity of wheat sold is between 5569·5 and 5570·5 maunds.

(*d*) The quantity of wheat sold is 5570 maunds correct to 0·008 percent.

Methods of approximation:—Approximation can be done (1) by discarding digits to arrive at a round number. In this no consideration is made of the digit which is more than half or less than half. It is simply dropped, thus:—

273,200 maunds would become 273,000,
594,800 maunds would become 594,000,
22.85 ounces would become 22, and
27.12 ounces would become 27.

(2) When approximation is done by using the next higher whole number the figure will be raised as shown below:—

273,200 maunds would become 274000
594,800 maunds would become 595000
22.85 ounces would become 23
and 27.12 ounces would become 28.

(3) The proper method is of approximating to the ***nearest*** whole number In this if the digit to be approximated is more than half it will be raised while one that is less will be reduced.

273,200 maunds would would become 273,000
594,800 maunds would become 595,000
22,85 ounces would become 23 and
27.12 ounces would become 27.

When approximation has been made it is necessary to bear in mind that in doing mathematical operations of the approximated figures the error is likely to be increased or decreased according to the nature of the operation. Really speaking an error in approximation does not refer to any mistake in compilation but to the difference between the actual and approximate figure.

Errors:—In statistics the word error is not constructed as mistake but as one denoting the difference between the estimation and the actual measurement. An error can be measured absolutely or relatively. An *absolute error* is the difference between the estimated and real value while a ***relative error*** is the ratio of the absolute error to the estimate. If the actual income of A is 18,160 and it has been estimated at Rs. 18,000 the absolute error is 160 and the relative error is $\left(\frac{160}{18,000}\right)$ i. e, 0.009 or 0. 9 percent For comparative purposes a relative error is a better guide than an absolute one. For instance, if the number of labourers in mills A and B were estimated at 2700 and 1000 respectively when actually the mill A has 2680 and B 960 the absolute error in case of A is 120 while in B it is 40. the relative error being 0.004 in A. and 0·04 in B. The relative error has thus revealed the difference more clearly.

Sources of statistical errors:—Errors of origin occur because of a lack of proper definitions of the subject matter or of a bias in information. While taking census, if proper definition of a married person is not given, some might class a widower or a widow under married while others will do otherwise. Errors of manipulation are due to arithmetical as well as clerical errors occurring involvntarily in measuring, counting or approximating figures. If samples taken are not sufficient, errors of inadequacy will spring up.

Biassed and unbiassed errors:—***Biassed or Cumulative*** errors which are based on prejudice err always on the same side or direction and thereby cumulate the error. ***Unbiassed, Compensating or Fortuitous*** errors on the contrary err on either side and so they tend to eliminate each other thereby not materially affecting the results. If a merchant measures cloth with a yard of

35 inches in every yard he will leave one inch and with the increase in the number of yards measured the mistake will be increased. On the contrary, if his yard be of correct measures but in measuring he makes mistakes, sometimes they will be of short measure while at other times they will be of longer measure, and so these mistakes would compensate against each other. In the former case, however, they would accumulate as would be seen from the following example of approximation,

Actual figure 1	Estimated figure by taking		
	Lower digit 2	Higher digit 3	Nearer digit 4
72.6	72	73	73
72.4	72	73	72
72.7	72	73	73
72.3	72	73	72
72.8	72	73	73
72.3	72	73	72
435.1	432	438	435

The actual total is 435.1 but in columns (2) and (3) the figures have been approximated with a bias. In the former they have been written at a lower digit while in the other they have been placed at a higher digit. The result is that in both the cases there has been a cumulative error. In the last column the approximation has been done without any bias or prejudice and the figures have been taken to the nearer digit it may be lower or higher. The total. therefore, has been found to be almost correct as the errors have been compensatory. In research work which depends mainly on the data collected it is very necessary to guard against errors and more particularly against those of biassed type.

Exercise

1. Distinguish between statistical description and statistical induction.
2. Give the advantages of a sample enquiry over a census.
3. 'Organized knowledge is representative in character', Discuss.
4. Explain the Law of Statistical Regularity.
5. Give the essential requisites of a good sample.
6. Give all the methods of sampling and state with reasons the method you recommend for an industrial survey of a district.
7. How will you test the reliability of the samples taken in a statistical investigation?
8. Mention the advantages of approximation in statistics.
9. Give the sources of statistical errors and state how they can be avoided.
10. What is random sampling ? Show its superiority from statistical standpoint over deliberate sampling.
11. Write short notes on:—
    Inertia of large numbers, stratified sampling, stability test, relative error, biassed errors.
12. What is a statistical error? How far does it generally affect the value of a statistical investigation?

CHAPTER IV

COLLECTION OF STATISTICAL DATA

After the preliminary factors such as defining the problem, scope of enquiry, degree of accuracy, requisites of good statistical data including sampling and possibility of likely errors have been discussed at length, we take up the problem of the collection of statistical data. As stated already this is the foundation of statistical investigation and upon its accuracy and thoroughness the quality of succeeding operations of tabulation, summarization and interpretation depend.

Primary and Secondary data:—Statistical data are usually classified on the basis of proximity to original observations as Primary and Secondary. Data gathered by original observations measurement, count or original recording during the investigation are known as *Primary data,* while data collected originally by some one else are known as *Secondary data* and so are removed one or more steps from the original enumeration. The data collected by the Census organisation are primary data for the Census office but are secondary data when used by others. It will be seen from above that this distinction between the two is not clear-cut but one of degree. Data which are primary in the hands of one become secondary in the hands of the other. In the same manner very often a distinction is made between primary and secondary sources of statistical data. The original compiler of the data is the primary source but one who gives the data originally gathered by some one else is a secondary one. Methods too are classed as primary and secondary. One collecting original data is classed as Primary while the other collecting data which have been already worked over is called Secondary method.

Methods of collecting data:—The different methods of collecting material for statistical purposes can be divided into two sections depending upon the class of data which may be primary or secondary.

A—**For the collection of primary data:**—

1. Direct personal investigation.

2. Indirect oral investigation.
3. Estimates from correspondents.
4. Schedules filled by informants.
5. Schedules in charge of enumerators.

B—**For collection of secondary data**

The main sources of secondary data can be divided into the following groups :—

1. Government publications.
2. Publications of commercial and trade associations.
3. Information obtained from newspapers and periodicals.
4. Private organisations.

1. **Direct personal investigation:**–This methods is particularly useful when only a few cases of intricate nature are to be studied intensively. The classic example of the type of investigation is of Le Play who spent a long period in studying Workingmen's budgets in Europe. He actually lived with some families for this purpose for days together. This method can be followed only if the cases to be investigated are few. Though not so intensive the work of Rowntree in York in 1900 was of a similar nature. No doubt by this method the results achieved will be far more accurate, practically because of the personal interest and supervision at every stage of the enquiry. What investigation would be more reliable than the one conducted with one's own eyes and ears. The chief drawback, however, is that an extensive field of enquiry cannot be covered by this method. In the first place in some cases such an enquiry may not be feasible and, in others, by the time it is finished that data might have gone out of date because of the considerably long time taken to collect it. As the number of cases that can be studied by this method is bound to be meagre they cannot be considered sufficient to constitute a fair sample of the statistical universe. Besides this, there is a great chance of personal bias influencing the investigation as it is likely that the investigator may start with some preconceived ideas of his own apart from the fact that such investigations are usually costly as well as dilatory.

In the event of the method being not practicable in some cases, as the investigator may not be in a position to be on the spot or to devote the necessary time, evidence can be collected by questioning and cross-examining persons who are directly in touch with the facts under investigation. This method will not, of course, be so successful as the investigator has to depend upon the knowledge as well as on goodwill of the other party. But with reasonable care, simple, short and clear-cut questions and courteous behaviour on the part of the investigator the results can be improved.

2. **Indirect oral investigation:**—This method is resorted to when the field of investigation is so vast that it cannot be conveniently covered by direct personal investigation or when the nature of information to be obtained is of complex nature and the informants are indifferent to give it if approached directly. Evidence is then sought from third parties who are in touch with the facts under enquiry. Enquiry committees and commissions adopt this method, but great care as well as discretion is necessary in recording such evidence. This type of investigation is done by keeping some standard set of questions at hand. It is, however, risky to rely wholly on this type of evidence unless it is, ascertained that the informant really possesses adequate knowledge of the facts and that he is not in any way prejudiced or motivated in presenting the facts in a particular way. Even with all sincerity due to optimistic or pessimistic nature of a person facts are likely to be presented differently.

3. **Estimates from Correspondents:**—Under this plan for investigation there is no formal collection of data but estimates are called for from local agents and correspondents and reliance is placed on their ability to obtain them. This method is particularly advantageous in conducting extensive investigations cheaply and expeditiously but where accuracy is not of *very* great importance. This method differs from the two other methods of getting questionnaries or schedules as in this method it is usually presumed that the correspondent does not possess very accurate knowledge and that the facts supplied are more estimated than counted. Various crop reports obtained by the governments are of this type. If individual reports are considered they may not be

accurate but taken together the results are likely to be fairly accurate as the errors involved in individual reports are likely to compensate each other.

4. **Schedules filled by informants:**—In this method schedules of questions called 'Questionnaires' or Blank forms (so called because of the provision of blank space for writing answers) are sent to informants with a request to return them duly completed. This method differs from the previous one only in that the questions expected to be answered are those about which the informant is supposed to possess definite as well as accurate knowledge. This method is inexpensive and, if honest co-operation and active support of the informants can be enlisted, the results are fairly accurate and reliable. But in the absence of any type of complusion or active interest a large percentage of schedules is not returned and those that are received are mostly from interested parties. Due to want of proper care many schedules returned are incomplete as well as inaccurate. If a letter couched in polite language with a stamped envelope is sent the response is likely to be better. This method is largely adopted by private individuals as well as by institutions and even by the government to collect statistics of wages, unemployment, factories, etc., as by this method the investigation work is cheaply as well as quickly done, particularly when a large territory is to be covered. It is but natural that accuracy by this plan of work will not be as high as is possible when the schedules are in charge of enumerators.

This method, however, is defective in some respects. Persons are usually willing to give information orally but many of them will not commit it in writing. Interested persons make it a point to return schedules duly completed but many others for no particular reason do not send them and so the results obtained are likely to be on the whole not of a representative character. To achieve better results and co-operation it is necessary to write polite letters to informants promising to treat the information supplied as strictly coufidential. Again, the questions referred to should be on matters with which the informant is expected to be fairly familiar, as it is unlikely that one would hunt up information to fill in the schedule. The questions should be clearly worded and should admit of no ambiguity. Their number too should be small and mostly relating to present day matters as to a great extent inaccuracy creeps in the answers requiring strain on memory. They should not give offence of any kind or create suspicion in the mind of the informant. If these precautions are taken and careful checking of the schedules is done we can achieve a fair amount of success by this method. To secure answers from many informants the usual follow-up plan adopted by commercial firms will be of considerable help.

5. **Schedules in charge of enumerators:**—When the schedules are to be filled up by informants the difficulty experienced is that many informants do not care to return them without even any particular reason. To get over this defect this method of sending enumerators to collect the necessary information is adopted and has been found by experience to yield better results. The success of this method depends largely upon the type of enumerator employed. As they are to personally get the schedules completed there are less chances of omissions or wrong interpretations of questions. If they are given some initial training before they are sent out, uniformity, proper selection of samples, etc, can be ensured. This method, though it yields good and correct result, is costly to be adopted by private individuals or by institutions and is therefore mainly employed for government work.

Enumerators:—It is the enumerator who is primarily responsible for the statistical investigation and as such the success or otherwise of the enquiry mainly depends upon him. As it is not possible for the person–in–charge of the investigation to approach the whole 'Universe' he takes the assistance of several enumerators. It is very necessary that they must be well-behaved, cool-headed, courteous, tactful, diligent as well as intelligent, besides being honest and sincere. As they have to get the information for writing schedules they have to be polite and courteous; otherwise the other party may not respond whole-heartedly. While dealing with illiterate persons they will have to be cool-headed as they are not likely to understand the questions clearly and hasty as well as short-tempered enumerators may not succeed in getting the required information. Great tact is required to elicit answers and sometimes to cross the party when he is trying to give vague or evasive answers. For instance, in the last census it was required to ascertain the age of a woman at the time of the birth of the first child. This information can be easily obtained by ascertaining the present age of the mother as well as that of the first issue. Enumeration is a tedious work and only diligent and sincere persons can do it well. Above all, as Dr. Bowley has said, 'In collection and tabulation common sense is the chief requisite and experience the chief teacher.' It has been often noticed that enumerators fill in questionnaries without actually approaching the parties or at least without ascertaining full details. In making statistical survey it is a good policy to have supervisors to check the work of enumerators employed as well as to guide them. Careful selection of enumerators as well as supervision over their work will to a great extent raise the efficiency of the investigation work.

Questionnaries:—For collecting statisiical data the information obtained is recorded on specially designed forms called 'Schedules' or 'Questionnaries.' Though both the words are often used without any distinction, 'Schedules' are forms which are ordinarily filled by trained enumerators who go about collecting the required information while questionnaries are fiilled dy informants. The success of the enquiry undertaken largely depends upon the drafting of a suitable questonnaire and so it is necessary that proper care should be taken in its preparation.

Requisite of a good schedule or questionnaire:—A well drafted schedule or questionnaire should as far as possible satisfy the following requirements :—

1. There should be as few questions as possible but not so few as not to be able to have full information and yet not so many as to tire out the patience of the informant.

2. Questions should be clearly worded making them easy to understand and free from ambiguity of any type. When a question about price is asked it should be clearly stated whether cost or market price is to be taken into account.

3. The questions should be such can be answered without any kind of bias and by 'Yes' or 'No' or by simple numbers.

Such answers simplify the work of tabulation. Replies like 'Fairly well' 'So so', 'plenty' etc. serve no useful purpose in statistical investigation and so where opinion is required the questions should be so worded as to require an unqualified affirmative or negative answer.

4. The questions should not be unnecessarily inquisitorial so as to offend the party concerned or to arouse antagonism in any way or such as would mean asking some most private or confidential matter of the informant.

5. As far as possible the questions framed should be corrobarotary at least on points of importance. They will serve as a good check on answers received.

6, The questions framed should be such as can be answered by the informant with definite knowledge and information at his command.

7. The question should be directly related to the point of information desired. They should be sufficiently inclusive but unnecessary information should not be solicited.

8. In case the questionnaire is to be filled by an informant it is necessary that a statement giving reasons for the questionnaire and the use to which it is to be put should be given with a request for co-operation.

Specimen questionnaires:—Two questionnaires, one issued for an enquiry into the 'After careers of graduates' instituted by the University of Bombay and the other for an enquiry into the working of Co-operative Housing Societies are given below;—

AN INQUIRY INTO THE AFTER CAREERS OF GRADUATES OF THE BOMBAY UNIVERSITY

1. Name ... ...

2. Sex ... ...

3. Age ... ...

4. Religion ... ...

5. Married, Unmarried, Widower or Widowed ... ...

6. University Degree or Degrees

taken in the year—

1935 ... ...
1936 ... ...
1937 ... ...

7. Whether employed or unemployed? ... ...

8. If employed, whether employment is permanent or temporary, part-time or whole-time ...

9. Means adopted to secure employment ... ...

0. Present occupation ... ...

11. Whether eneaged in a business or profession on your own account or in service ... ...

12. If in service, state whether employed ... ...
(*i*) by the Bombay Government,
(*ii*) by the Government of India,

(*iii*) by railways, semi—Government concerns like the Bombay Municipality, Bombay Port Trust, etc.

(*iv*) in private schools,

(*v*) in commercial offices, Banks, etc. ... ...

13. If engaged in a profession please state which ... ...

14. Are you following the profession because of your failure to get an appointment ? ...

15. What is your starting pay? ...

16. How long after taking your first degree and wishing to secure permanent employment you secured (*a*) temporary employment, (*b*) permanent employment. ... ...

17. If employed temporarily, please state the total duration of employment since taking your first degree ... ...

Note—If engaged as a private tutor this should be specifically mentioned.

CO-OPERATIVE HOUSING SOCIETIES

Province......

1. Number of Societies...
2. Type...
   (a) Co-partnership...
   (b) Tenancy ownership...
3. Total membership...
4. Total share capital...
5. Methods of financing used—
   (a) Government loan...
   (b) Deposits from members ...
   (c) Deposits from non-members
   (d) Debentures.........
   (e) Loans from charitable institutions...........
   (f) Other sources............
6. Average rate of interest at which funds are obtained.........
7. Rate of interest charged on loans to members...
8. Area of the Co-operative Housing Societies......
9. Year in which the first society was registered.........
10. Number of plots and their average size......
11. Can a member hold more than one plot?
12. How are the common expenses shared?
13. Is there any help given to societies by–
    (a) Municipality–
    (b) Government–
    (c) Any other body–
14. Any other feature.........

Secondary data:–The collection of primary data is costly as well as laborious and so we have in many cases to fall upon secondary data. It is true that the results obtained from this source are not as satisfactory as those from the former. Besides this the statistics already collected may or may not be in conformity with the requirements of our present enquiry, as terms may have been used in peculiar senses or measurings may have been imperceptibly changed. In short, as Connor has said, 'Statistics, especially other peoples' statistics, are full of pitfalls for the user unless used with caution.'

Scrutiny of secondary data:—When statistical data is derived from the secondary source it is likely that due to limitations or inaccuracies present therein the conclusions drawn therefrom without necessary adjustments may not be correct. It is, therefore, incumbent to scrutinize the data carefully in light of the following points:–

1. The status of the compiler as well as his ability and means to collect statistics should be taken into account. A government department or a statistician of standing can be relied upon but data collected by inexperienced or interested parties will have to be taken with proper secrutiny.

2. The source of information is equally important. Statistics collected by municipalities and local bodies for the government departments are likely to be more reliable than the material obtained from private parties.

3. The scope and object of enquiry of the original investigation will have to be looked to with caution as if the objects of the two enquiries are not identical, the data already collected may not serve the required purpose.

4. The method of collection of data plays an important part. There is a difference between the data collected by census and sample methods and between the data collected by random and deliberate sampling. The statistical matter collected by personal enquiry is certainly far more accurate than that obtained with the help of questionnaires.

5. The definitions of different items should be seen very carefully. The exact significance of the terms used should be

observed minutely. A house may mean a dwelling with one or more rooms, so also profits may mean gross or net.

6. In considering statistical matter the degree of accuracy observed in collecting the data under consideration will have to be seen in determining the accuracy of the conclusions to be drawn therefrom.

7. Very often it happens that the conditions may not have remained homogeneous, Tastes, habits and customes might have changed after the data was collected and the material may not properly fit in the changed circumstances. Some years ago by transport was meant only water and land transport but in the present times air transport has to be included. Similarly, geographical divisions might have been changed due to political circumstances.

8. The period covering the data particularly in the case of time series is of importance, as the data for a short period is likely to be influenced by short-term fluctuations.

9. We must also enquire about the organization that supplies the data with a view to noting whether there is any possibility of bias or prejudice influencing the material collected.

In short we have to see from various angles whether the data available can be suitably applied to the problem in hand. It is no doubt true that this method of collecting data is far cheaper but cheapness alone cannot be the factor to be taken into account.

Important sources of secondary data:—As said already the important sources of secondary data are as under:—

1. **Government publications:**—Every government in the course of its administration collects statistical material. Various returns regarding crop statistics, vital statistics, trade statistics etc., are called for and unless the data be of a confidential nature, people have access to the material which may or may not exactly fit the enquiry on hand. Besides these, now and then special enquiry committees are appointed and their reports are published and

made available to the public. When this material is to be made use of in addition to other things particular care should be taken to study its limitations and the purpose for which it was collected.

2. **Publications of commercial and trade associations**:– The various chambers of commerce and trade associations of different types issue their reports on special topics and also conduct enquiries in certain cases. This material as well as trade papers are of great value for collecting statistical material.

3. **Information obtained from news-papers and periodicals**:—'Capital', 'Commerce', 'Eastern Economist', 'Indian Finance,' and other periodicals give considerable amount of statistics in their issues. Besides, some of them publish year books giving consolidated statistical material on various topics. Statesman's Year Book, Times of India Year Book, Indian Finance Year Book can be cited as important sources of statistical information on commercial as well as on other topics.

4. **Private organizations**:–Besides these there are various other clubs and research bureau that collect statistics. University departments carry on research in different topics and statistical material is collected which is available for use to others as well.

In addition there are unpublished statistics but in all cases it is very imperative that proper precrutions should be taken as discussed already to see how far they can be used without necessary adjustments to suit the requirements of the problem in hand.

Choice of Method:—We have discussed at length the various methods commonly made use of in the collection of statistical material. Which particular method should be availed of will depend largely on the circumstances of each enquiry, particularly the amount of funds, man power, time required, accuracy desired, sources of information and the agency available. In some cases there is hardly any choice left. For instance, if a decennial census is to be taken, the government will have to be approached, as it is beyond the scope of any private individual or institution to conduct it with the required efficiency.

Exercises

1. Distinguish between primary and secondary data and state to what extent they may be considered reliable.

2. Give the important methods of collecting statistical data.

3. In which case is the method of direct personal investigation particularly useful?

4. The method of getting schedules filled by informants is cheap as well as good. Do you agree with this statement ? Give reasons in support of your opinion.

5. What qualities should a good enumerator possess ?

6. What precautions are needed in framing a questionnaire?

7. How will you test the reliability of secondary data ?

8. Give the important sources of secondary data and mention their limitations.

9. Write short notes on :—

   Schedule, enumerator, estimates from correspondents, Government publications.

10. Draft suitable questionnaires for enquiries regarding :

    (a) Effect of decontrol.
    (b) Sugar industry in Bombay State.
    (c) Cost of living of workers in a local factory,
    (d) Market research regarding the sale of radio in your town.

11. 'In collection and tabulation common sense is the chief requisite and experience the chief teacher.' Discuss this statement of Dr. Bowley.

12. Draft a suitable questionnaire for the industrial survey of a province.

CHAPTER V

CLASSIFICATION AND TABULATION

Statistical data collected in the course of a statistical enquiry very seldom comprises of simple matter which can be easily presented and understood without any special treatment. It is very difficult to appreciate at a glance or even after a careful study to bear in mind the information contained in the several schedules collected by enumerators. It is, therefore, necessary to condense the data by curtailing the amount of insignificant features and by bringing it into a form by which the significant details become more prominent. This is achieved by the process of Classification and Tabulation. The former consists in grouping the items according to their common characteristics and the later in arranging them suitably in columns and rows. This task appears at first sight to be very simple and requiring no trouble or thinking but one realizes only by experience the difficulties and pit-falls that are to be got over by exercising great care and patience in getting the mass of matter classified and tabulated.

Classification

The first and foremost step after the collection of the data is classification. For instance, when the census is taken a mass of matter in the form of schedules will be collected giving information about age, sex, civil condition and education. But it will need proper classification before this statistical data will serve any useful purpose. Groups will have to be made by taking into account common characteristics or attributes of the various units under consideration such as age, sex, etc. Classification is thus the process of arranging the available matter in groups or classes according to resemblance and likeness.

The main *object* of this process is to bring out points of similarity as well as of dissimilarity existing in various units and thereby curtailing unnecessary details and making it easier to form mental pictures of objects of perception and conception and also to save mental effort as well as energy and time. In addition, when the matter is available in a classified form, comparisons can be made and inferences drawn with more ease as well as accuracy. It also helps in bringing to light or at least in prominence that information which would have otherwise remained unnoticed because of its lying in a mass of matter.

Statistical data:—Before the information collected for statistical purposes is classified it should be critically examined and schedules found incomplete or those containing errors, inconsistencies and omissions should be corrected or rejected with a view to avoiding results being vitiated by the inclusion of schedules with these discrepancies. It is the weakest link in the chain that determines its strength, and even though the rejection of some matter will mean reduction in the statistical matter collected, this will certainly improve the quality of the work. In examining the matter it should be borne in mind that nothing is a *statistical fact* which does not correctly represent the whole data and has not been properly checked and verified. Taking a few abnormal cases here and there and generalizing therefrom is certainly dangerous. Complete information relating to the particular fact has to be collected and it should in no way be subject to personal bias or embellished by personal opinion.

Basis of Classification:—Statistical facts after they are properly scrutinized are classified according to different elements involved therein e. g. the Time element, the Space element and the Attributes. In doing so, all the statistical observations called *Variates* are considered similar in all respects except in the particular element under consideration. When they are classified according to the *Time*, *Temporal* or *Historical* element the essential characteristic seen is chronology, and so importance is attached only to the time at which the objects in question were measured or to the occurance of the events in question. In case of the Space element the variates are taken to be similar in all respects except their location in space. The element of time is considered insignificant as the observations are made at the same time. When classification of rainfall is given by time the rainfall of different years is shown but in case of spatial classification the recorded rainfall of different places will be given.

When the classification is on the basis of attributes all variates are considered as similar except for the possession or destitution of certain characteristics or attributes such as weights, height, salaries, etc. This group can be further subdivided into three sub-groups: (*a*) groups in which the attribute is either present or absent, *e. g.* a male or female; (*b*) groups in which the attribute is present in varying degree,· *e g.* groups showing the ages of persons and (*c*) groups in which the attribute is of varying quality. *e. g.* groups of persons with different professions such as doctors, pleaders, teachers, etc.

Simple and Manifold Classification;—Classification is also determined by two kinds of characteristics possessed by the individual units of a group: descriptive and numerical. The latter is susceptible to quantitive measurement such as age wages, heights, etc. and observations of this type can be classified according to class-intervals. Observations possessing descriptive characteristics can be classified by some natural or physical lines of demarcation such as sex, civil condition, literary, occupation, etc. When one particular characteristic is seen, say married or unmarried, literate or illiterate, only two classes are formed—one having the particular attribute and the other without it. The classification of this type where there are only two sub-classes is known as *Simple Classification or Classification by Dichotomy.* If we study more attributes at a time several classes are formed. If civil condition, literacy, as well as sex are studied at one time there will be several classes such as married literate women, unmarried illiterate males, and married illiterate women. When each class is thus divided into more than two sub-classes the classification is said to be *Manifold.*

Series and Groups:—The terms *series* and *groups* are often indiscriminately applied to the classifications described already. Strictly speaking, the term series conveys the idea of time and as such it is to be used in Historical or Temporal classification reserving the other term for the remaining two. A series, therefore, consists of a number of observations in which the time factor is predominant while in the case of a group the observations are spatial or attributive in character.

The groups can further be sub-divided as objective and subjective. If the observations in the group are of the same object they are called *objective* while the groups having observations of different objects are known as *subjective.* This distinction is to be borne in mind as the conclusions drawn from the objective observations are not likely to be applicable to those of subjective nature.

Seriation:—A set of statistical items arranged with respect to definite rules of classification is termed a statistical series. It is formed by arranging two variable quantities side by side so that measurable differences in the one correspond with those in the other and this process is known as *seriation.* The rainfall in Bombay in successive years will form a statistical series as the rainfall being one measurable quantity will be arranged against another measurable quantity i. e. different years. Similarly, the numbers of labourers in mills can be arrayed against the wages paid to them to form another statistical series. But in some cases it is not possible to form such a series. If we have number of deaths in some places or acreage under different crops in some fields it is not possible to form a series as only one and not two are variable quantities.

Continuous and Discrete Series:–Series formed as explained above can be continuous or discrete. In the former type the data are not capable of exact measurement or exact measurements are not available or the items pass from value to value with fractional difference, *e. g.* ages or heights of children. In the case of broken, discontinuous or discrete series the data is capable of exact measurement and between the values of the two succeeding items definite breaks or gaps are visible, as in the case of the number of workers in different establishments or number ef houses in a city, as these are incapable of subdivision to avoid breaks in the values of the two succeeding items.

Tabulation

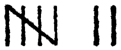
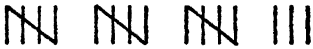
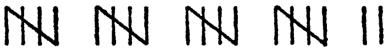
From the foregoing pages it is seen that classification is the process of arranging variates in groups or classes according to their resemblances thus giving expression to the unity of attributes subsisting amongst a diversity of individuals. Tabulation is the next step and consists in arranging the data in an orderly manner into (horizontal) rows and (vertical) columns that can be conveniently read in both the directions. According to Prof. Bowley it is an intermediate process between the accummulation of data in whatever form they are obtained and the final reasoned account of the results shown by statistics. It is a summarization process and involves an orderly and systematic presentation of the numerical data by re-arranging it with a view to elucidating the problem in hand.

Stages in Tabulation:—When schedules are submitted investigators they need classification, transcription, sorting and counting before they can be tabulated with a view to presenting the necessary information in a proper form. This process, therefore, can be split up in three stages. Schedules usually contain matter regarding various topics and therefrom each item giving the necessary information is to be extracted and then entered in slips or working sheets ruled specially for the particular problem. For transcribing this information from the original record to the secondary forms cards and tally sheets are made use of, particularly the former when mechanical methods are to be used. The second stage is of summarizing working sheets and entering the results in forms used for actual study which serve as basis for further manipulation, if necessary. Lastly, the final tables are written out entering therein the necessary details discarding other figures which were noted for making further calculations such as ratio and percentages as they are no longer necessary for explaining the final results.

Working Sheets:—In the first stage of tabulation working sheets or cards are written from schedules. In drafting the ruling for these sheets, particularly in the case of cards to be used for mechanical tabulation, great care and judgment is necessary. Sheets specially prepared should bear appropriate title and the columns thereon should have proper explanatory headings to avoid confusion. To guard against their loss or misplacing they should bear consecutive numbers. In the forms sufficient space should be kept for making additions as well as corrections which as a rule should as far as possible be minimised. To avoid confusion as well as mistakes too much matter should not be entered at once and there should be an arrangement for control columns as well as for check and cross-casts at each stage.

When the items to be counted are many the device of putting dots or ticks as shown below should be adopted to save time and space.

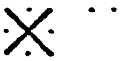
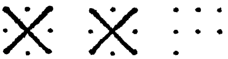
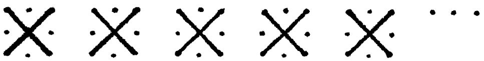
Crossing Fives

Successful candidates	Number	Total
CLASS I		7
CLASS II		18
CLASS III		22
Total		47

As after puttting in four ticks the fifth one is used for crossing this method is also known as *Four and Cross* method.

This can as well be done by putting dots as : · : : · : in place of ticks or strokes. In case the number of observations is large it is better to use *tally sheets* of crossing tens.

Crossings Tens

Successful candidates	Number	Total
CLASS I		12
CLASS II		27
CLASS III		53
Total		92

Spot diagram:—For compiling a frequency table observations are counted by putting dots or spots as shown in the above table. A diagram thus prepared is *sometimes* called a Spot Diagram.

Primary and Derivative Tables

Statistical tables may be classified under two general headings. (1) The Primary, Informative or Classifying tables are those in which data is presented substantially in the same form in which it was collected. They are merely convenient means of compiling and preserving data in a form of easy reference. In such tables no effort is made to present comparisons or relationships. (2) The Derivative or Interpretative tables are not the original tables but they are prepared after being at least partly subjected to one or more processes of grouping, totalling, averaging, etc.

Simple and Complex Tables

Statistical tables can also be classified as simple and complex.

(a) **Simple and Single Tabulation:**—In such tabulation information is given in respect of one or more independent questions having no direct connection amongst themselves. Such a table is also known as a *Table of the First Order.*

Financial position of Indian Ports in 1944–45

(in lakhs of rupees)

Port	Income	Expenditure	Capital Debt
Calcutta	703·1	569·4	2011·5
Bombay	474·6	384·1	1581·6
Madras	103·1	93·8	129·9
Karachi	162·4	118·2	235·7
Chittagong	7·0	9·2	53·4*

*This figure includes a loan Rs. 10·5 lakhs from the Government for port improvement.

(b) **Complex Tabulation:**—In this type several aspects of the subject matter under review or analysis of totals are shown and answers given have connection among themselves which is not the case in the case of a single tabulation.

Population in 1941. (*in thousands.*)

Province	Male	Female	Total
Bombay	10817	10033	20850
Madras	24557	24785	49342
Bengal	31748	28559	60307
Total	67122	63377	130499

In the above table the sub-division of the total population according to the two categories (e. g. population according to sex as well as provinces) has been given and answers are thus available to two mutually dependent questions. As the sub-division is into two distinct categories this tabulation is known as *Double Tabulation* or *Table of the Second Order.* In the following table the sub-division is based on three distinct categories of age, sex and locality and so this is known as *Treble Tabulation* or *Table of the Third Order.*

Distribution of population by age, sex and locality.

Age	Urban			Rural			Total		
	Male	Female	Total	Male	Female	Total	Male	Female	Total
Below 10									
10-25									
25-30									
50-75									
75-									
All ages.									

Distribution of population by locality, caste, and civil condition.

Locality.	Caste.	Age.	Male.			Female			Total		
			Married	Un-married.	Total.	Married	Un married.	Total	Married.	Un-married.	Total.
Urban	Hindus.	0–30									
		30–60									
		60–									
		Total.									
	Non-Hindus	0–30									
		30–60									
		60–									
		Total.									
Rural.	Hindus.	0–30									
		30–60									
		60–									
		Total.									
	Non-Hindus	0–30									
		30–60									
		60–									
		Total.									
		Grand total.									

In the same manner it is possible, though with difficulty, to have tables formed to show more than three categories and then it is known as ***Manifold Tabulation.*** In the table given on page 57 four categories i. e. locality, caste, age and civil condition have been shown. No doubt, because of the inclusion of so many items, the table is not very handy as well as easy to follow.

Arrangement of tabular facts:—Tabular facts may be arranged in the form of a Simple or a Frequency table to suit the requirements of the particular problem.

(1) Simple Table:—A simple table may be prepared for recording information applicable to groups or series. In the case of groups, tables can be arranged mainly in three ways: (a) according to alphabetical order, (b) according to the order of magnitude and (c) according to their geographical distribution.

Number of factory workers in British India in 1944.

Class of factories	Average daily number
Textiles	993,269
Engineering	265,392
Minerals and Metals	91,126
Food, Drink and Tobacco	132,384
Chemicals and Dyes	88,813
Paper and Printing	52,696
Wood, Stone and Glass	96,189
Gins and Presses	14,850
Skins and Hides	34,624
Miscellaneous	35,477
Total	18,04,820

The above is the usual form of a Table but it can be rearranged according to the alphabetical order or according to the order of magnitude. If arranged by the former method it will appear as under :—

Number of factory workers in British India in 1944.

Class of factories	Average daily number
Chemicals and Dyes	88,813
Engineering	265,392
Food, Drink and Tobacco	132,384
Gins and Presses	14,850
Minerals and Metals	91,126
Miscellaneous	35,477
Paper and Printing	52,696
Skins and Hides	34,624
Textiles	993,269
Wool, Stone and Glass	96,189
Total	18,04,820

In case it is to be written according to the order of magnitude the number employed in textile factories being the highest will have to be placed first in the list and will be followed by Engineering, Food, Drink and Tobacco, Minerals and Metals, etc.

The alphabetical arrangement is no doubt very useful for reference purposes but for interpretation and analysis it is of little use. The other method, though not suitable for reference purpose, is of great value in arresting the attention of the readers to the matter of importance.

The geographical arrangement is made on the distribution of the variates on the territorial basis. This method is particularly useful in some cases as the characteristics of variates in adjoining places are likely to be more similar and in making comparisons and drawing conclusions this factor is found to be of great value. Tables giving population according to provinces or exports or imports given according to countries are examples of this type.

In the case of *series* it can be arranged only in the chronological order as the main object of such a table is to show the developments over a period of time. Usually the items are arranged from the earliest to the latest times but some prefer the reverse order with a view to giveing more importance to the events in the present times.

Frequency Tables:—We find by experience that all objects are not of the same type and they vary in size, shape, etc. In statistical parlance a *variate* is a thing observed and its value is known as *Variate value.* Variates may be Simple or Compound or derived, the latter type being obtained by striking some sort of relation between two or more variables already recorded, e. g. ratios, percentages, etc. A *Variable* is anything that exhibits differences in magnitude or number and the values of which can be measured in appropriate units. The variables are of two types: Continuous and Discontinuous. A *continuous variable* is one which is capable of taking all possible values between a certain limit while a *discontinuous or discrete* variable is one which is incapable of taking all possible values and thus there are gaps in the values of such variables. If heights or weights of different persons are taken there is a possibility of having all figures of heights and weights in certain limits, e. g. 5′–5″, 5′–5·1″, 5′–5·2″ but if the number of persons employed in a factory is taken it can be only in integers and so naturally there will be gaps and so this variable is classed as discrete.

When this variation is due to lapse of time, such as imports in various years, a time series is formed but when it is at the same time such as heights of different persons recorded at a particular time it is known as synchronous variation and this is at the basis of frequency tabulation. *A frequency table* is framed by recording the frequency of occurence of items i. e., the number of times a particular value of a variable occurs in a set of observations. As a matter of convenience all the observations are distributed in different classes, groups or grades and the manner in which it is done is known as *Frequency Distribution of Variables.*

Marks of 100 students

Marks	Frequency of	Cumulative frequency of
0–10	4	4
10–20	7	11
20–30	10	21
30–40	15	36
40–50	8	44
50–60	24	68
60–70	13	81
70–80	12	93
80–90	5	98
90–100	2	100
Total	100	

This table has beeen compiled from the mark sheets showing marks of 100 candidates arranged according to their roll numbers.

7 20 95 31 51 63 52 32 81 28 80 53 62 40 63 45 73
10 61 9 62 50 70 27 71 47 64 34 65 22 72 57 74 58
22 33 52 23 5 33 61 53 37 66 42 54 57 12 55 26 58
74 48 60 35 79 52 34 3 14 25 56 37 76 36 38 75 59
24 68 13 78 35 54 72 58 55 16 67 57 56 83 44 54 46
56 91 43 29 69 55 36 87 53 38 77 17 82 18 39.

From the foregoing table it will be seen that all the hundred observations have been divided into 10 groups called *classcs* and the distance between the limits of two classes is known as *class interval* which in this case is 10.

Class interval:—As far as possible when making groups or classes the class-intervals taken should be equal. If otherwise, it is likely that many readers may assume them to be equal without scanning them carefully and thus forming an erroneous impression. It is not easy in statistical operations to deal with classes of unequal sizes and, again very often, there is a possibility of making unequal classes for some ulterior motive. In addition, in determining a class interval it is necessary to see that the number chosen should not be too small or too large. In the case of the former there will be several classes, some of them having a zero frequency and the advantage of summarization will be lost. In the case of the latter there will be very few classes. The size of the class-interval should be such as would show the characteristic features of the distribution. As a general rule, it is better to ascertain the range between the values of the lowest and highest variaties in rank and to divide it by 12. This will give an approximate figure to fix the class interval. It should, however, be a round figure and usually figures like 1, 3, 5, 7, 10, 15, etc. are preferred, Ordinarily the number of classes made should not be less than 10 to 12 and not more than 25 to 30. In short, the underlying idea as stated by G. U. Yule is that we should be able to treat all the values assigned to any class without serious error as if they were equal to the mid-value of the class interval.

Class limits:—The class limits can be shown in one of the following ways ;—

(a)	(b)	(c)	(d)
0–10	0–9.9	0 and under 10	5
10–20	10–19.9	10 and under 20	15
20–30	20–29.9	20 and under 30	25

In the first method there is the overlapping of figures. 10 occurs in the first as well as in the second group and 20 is in the second as well as in the third. This leads to ambiguity though the practice is to take figures upto 10 in the first group and 10 and upwards but upto 20 in the second. This difficulty does not arise in the second method and, in addition, it affords an advantage of exhibiting the degree of accuracy adopted in classification. In the third method the meaning is better expressed but the actual class limits are not very clearly understood. Though the second method is better in practice the first is often used. It is not uncommon to find tables with mid-points only as shown in the last column of the table given above.

Cumulative Frequency Tables:—Frequency tables are of two types Simple and Cumulative. In a ***Simple Frequency Table*** groups of observations are divided into smaller and more homogeneous classes and the numbers of items to be found in each class are given. In a ***Cumulative Frequency Table*** class frequencies, instead of being shown separately, are totalled so as to show the total frequency at different stages of the table and so the frequency of each successive class goes on increasing as it includes the frequencies of all the previous classes. Such tables are very useful for making comparisons and for graphical work. In the foregoing table giving marks of 100 students in the last column the cumulative frequency has been given but in such a table the way of writing the classes is different as will be seen from the following table :—

Marks of 100 students

Marks	Number
Under 10	4
,, 20	11
,, 30	21
,, 40	36
,, 50	44
,, 60	68
,, 70	81
,, 80	93
,, 90	98
,, 100	100

In the above table the classes made are of under i.e. less than 10 marks, less than 20 marks, etc. and therefore such a table is sometimes known as '*Less than Table.*' *A* '*More than Table*' can be prepared from the same data and will be as under :—

Marks of 100 students

Marks	Number
Above 100	0
,, 90	2
,, 80	7
,, 70	19
,, 60	32
,, 50	56
,, 40	64
,, 30	79
,, 20	89
,, 10	96
,, 0	100

Advantages of frequency tables:—Frequency tables are useful in various ways. They help in minimising the errors arising out of biassed observations. As the figures will be grouped the errors are likely to be compensated against each other. Due to grouping the number of items for calculations are considerably reduced. Comparisons can be easily made and there is a great possibility of some tendencies ordinarily invisible being revealed. Above all, extensive information is made available in comparatively limited space which means saving of time as well as energy without much loss of efficiency.

Rules for Tabulation:—The utility of tabulation depends upon the manner in which it has been designed. Though it is difficult to lay down comprehensive rules the following will be of great use in constructing good statistical tables:—

1. Every table should bear a suitable, self-explanatory, comprehensive and yet, brief heading. When the lengthy title is unavoidable it should be split up into phrases and written in separate lines. For each column there should be a heading, called *caption* and for sub-columns *sub-captions* and they should be written as far as possible horizontally. There should as well be titles, known as stubs, at the left of each row.

2. The columnar layout of a table should be carefully thought out and it should not be too complicated or confusing. If possible different types of lines thin, thick, double etc. should be drawn for distinction and easy reading of the table.

3. The size of a table should be kept as far as possible handy as an unduly long table is difficult to follow carefully and by folding it will soon be spoiled and cut into small pieces. In case it is difficult to shorten the size of the table it should, if possible, be split into convenient parts.

4. Economy of space should wherever possible be done. Unimportant matter should be dropped and large numbers should be curtailed by dropping some digits but this should be shown in the heading, as '000 omitted.'

5. Wherever necessary explanatory foot-notes, sources and references should be given to avoid ambiguity.

6. A rough table should always be drafted before one is ruled and figures entered therein. This will give an idea of the necessary space and avoid errors and omissions at a stage when it is not easy to rectify them.

7. The arrangement of the table should be logical and items related to each other should be placed near about and if possible in groups. Derivative figures such as total, averages, and percentages should be placed near their originals. Columns and rows should be numbered for identification as for giving reference it is easier to quote numbers rather than the title of the column.

8. To minimise errors provision for checks and counter checks should be made in the tables. Errors and scratching should be avoided and if at all it is very necessary it should be very carefully and legibly done.

9. It should be noted that the cost of tabulation and labour involved should not be very heavy and in no case disproportionate to the value of the results.

In short, in the work of tabulation the *essential requisites* of a good statistical table, i. e. comprehensiveness, ease of comparability, referencing, compactness and cost should be the guiding factors.

Method of giving emphasis:—Sometimes it is found necessary to lay emphasis on some columns or on some figures of the group. This can be achieved in one or more of the following ways;—

(a) By specially marking such items by underlining or by drawing a circle, round the entry.

(b) By using different types of letters if in print, e. g. Italic or heavy type, or by writing the matter in different colours.

(c) By entering first the matter to be emphasised. If magnitude is to be emphasised the table will be arrayed in the order of magnitude.

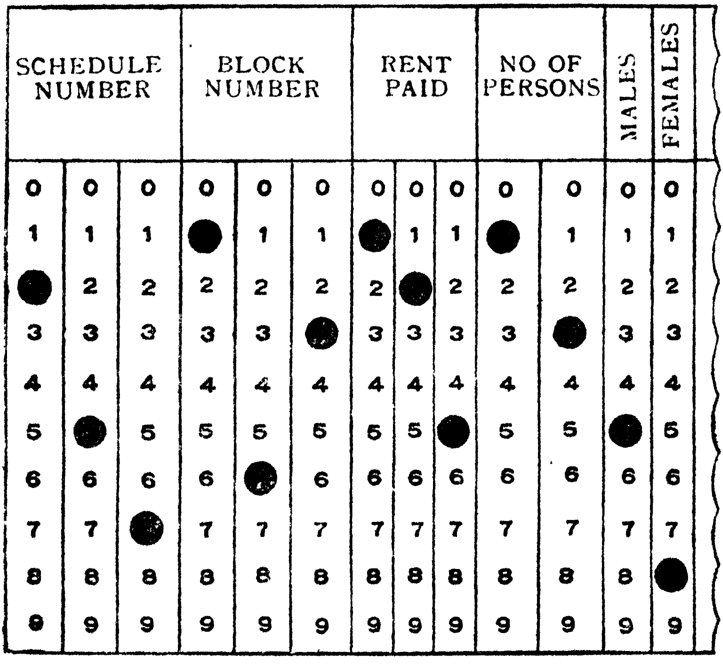
(d) If the space is sufficient it can be done by leaving more space between the groups to be emphasised.

(e) By using distinctive lines such as single and double or light and heavy between different sets of figures. This will give prominence to the desired matter.

In addition to this, percentages, ratios, averages and estimated figures as well as negative questions can be shown in red ink in hand-written tables and in Italic types in printed tables to lay emphasis on them.

Mechanical Aids in Tabulation:–Statistical work is every day attaining greater importance and so for classification and tabulation mechanical aids are being used more and more day by day. They are particularly useful for punching, sorting, counting and calculation work. The work done is very accurate and the speed with which the various operations are carried out is surprising. Cards can be sorted at the rate of 60,000 per hour for which several hours will be necessary if done otherwise. Due to this high speed in times of emergency statistical work can be completed in a very short period. Taking all these points into account the cost is not comparatively very high.

For using mechanical aids *Punch Cards* are required to be specially prepared. In these cards ***Fields*** or columns are made for recording each item. After punching holes in the fields according to the data the cards are mechanically sorted and counted to have the necessary data for tabulation.



SCHEDULE NUMBER			BLOCK NUMBER			RENT PAID			NO OF PERSONS		MALES	FEMALES
0	0	0	0	0	0	0	0	0	0	0	0	0
1	1	1	●	1	1	●	1	1	●	1	1	1
●	2	2	2	2	2	2	●	2	2	2	2	2
3	3	3	3	3	●	3	3	3	3	●	3	3
4	4	4	4	4	4	4	4	4	4	4	4	4
5	●	5	5	5	5	5	5	●	5	5	●	5
6	6	6	6	●	6	6	6	6	6	6	6	6
7	7	●	7	7	7	7	7	7	7	7	7	7
8	8	8	8	8	8	8	8	8	8	8	8	●
9	9	9	9	9	9	9	9	9	9	9	9	9

Fig.—A portion of a punch card

On this portion of the punch card the following information has been recorded by means of punching.

Schedule number	257
Block number	163
Rent paid	125
Number of persons	13
Males	5
Females	8

Exercises

1. Explain the purpose of classification and tabulation of statistical data. (1938–Oct.)

2. Describe the important bases of classification in the case of statistical data.

3. Distinguish between the terms 'series' and 'groups'.

4. Give the different types of statistical tables and give an example of a treble tabulation.

5. Rule a blank table to show the distribution of population in India according to sex and literacy.

6. Differentiate between

   (a) Continuous and discrete variables,
   (b) Ordinary and cumulative frequencies, and
   (c) More than and Less than tables.

7. Describe what considerations are to guide you in fixing the range, the class-intervel and ends of class-intervals for a frequency distribution. (1938–Oct.)

8. What is a statistical table? Mention the points about which you should be particular at the time of constructing a statistical table.

9. In what way emphasis can be given to certain items in tabulation and checks can be imposed to verify accuracy in tabulation?

10. What mechanical aids can be used in tabulation? Describe the use of punch cards in tabulation.

11. Write short notes on :—
Working sheets classification by dichotomy, discrete series, stubs and crossing tens.

12. The following are the rents in rupees of 60 houses and you are asked to tabulate the same by grouping them by intervals of 5 rupees.

38, 42, 49, 37, 82, 74, 75, 62, 54, 79
84, 57, 36, 41. 47, 44, 63, 69, 45, 84
47, 39, 48, 54, 61, 71, 72, 83, 80, 51
31, 38, 43, 57, 67, 77, 74, 64, 84, 81
79, 31, 39, 47, 53, 64, 73, 81, 39, 43
41, 42, 54, 59, 69, 75, 39, 85, 71, 63

13. Wages of 70 workers in rupees are given below. Tabulate by taking 10 as the class interval.

30, 45, 55, 65, 60, 90, 115, 85, 95, 110
95, 65, 75, 85, 125, 110, 87, 65, 100, 115
65, 60, 75, 95, 130, 95, 125, 115, 65, 70
95, 85, 65, 60, 80, 85, 75, 95, 55, 45
35, 45, 40, 85, 135, 145, 95, 65, 45, 25
115, 90, 80, 125, 130, 25, 90, 100, 95, 85
85, 120, 115, 40, 35, 35, 125, 105, 75, 45

From the above data prepare a cumulative frequency table.

14. The following are the marks of candidates who had appeared at the Efficiency examination. You are asked to tabulate them by taking a class interval of 10.

72, 19, 21, 32, 75, 68, 40, 52, 84, 18, 37, 12, 11, 27, 43,
18, 24, 11, 27, 39, 52, 93, 11, 7, 41, 19, 12, 23, 64, 52,
93. 99, 79, 84, 12, 11, 17, 19, 21, 29, 43, 33, 23, 94, 82,
27, 32, 12, 34, 54, 64, 74, 24, 12, 18, 39, 49, 77, 27, 17,
34, 84, 37, 77, 67, 57, 47, 37, 36, 35, 97, 34, 54, 91, 90,
70, 80, 60, 50, 40.

CHAPTER VI

AVERAGES

The term 'average' in the ordinary parlance is used to signify the type usually found but in Statistics the ' type ' which is common is known as 'Mode'. It is also used to denote the Arithmetic Average or Mean. If the heights of five persons are 62, 65, 64, 67 and 62 inches the average is ascertained by totalling heights of all the persons and dividing the sum by the number of persons; e. g. $\frac{62+65+64+67+62}{5}=\frac{320}{5}=64$ inches. Thirdly. it is also used as a generic term including the different kinds of averages e. g. Mean, Mode, etc. In statistics it means '*a simple expression showing the net result of a series or group*' and is thus a term which will give a fairly adequate idea of the group. For example, if we are engaged in studying the earnings of persons engaged in two establishments one having 3000 persons and the other 2000 it is difficult to have even an idea of the whole data by looking at the two lists giving the earnings of the various persons of the two establishments and the question of comparing them will be still more complicated. This difficult task is simplified by ascertaining for each series a *significant single expression* representing the whole series and thus conveying a fairly adequate idea of the whole group. The summarizing expression of this nature is known as an '*Average.*' The following are the forms of averages that are commonly used :—

(1) The Arithmetic Average also called the Arithmetic Mean, the Mean, the Common Average or simply Average.
(2) The Median.
(3) The Mode.
(4) The Geometric Mean.
(5) The Harmonic Mean.

Of these the first three are in general use while the last two are used only in special cases.

The Arithmetic Average

The Mean or the Arthmetic Average is the sum of values of the items taken and divided by the total number. Of all the types of averages this is commonly used and is hence called 'Common Average' or even simply 'Average'. Efforts have been made by writers like Dr. Bertillon and Dr. Venn to distinguish between 'Average' and 'Mean' but so far no general agreement has been reached as regards the exact terminology. Dr. Bowley has suggested that the distinction should be made by taking 'average' as a purely arithmetic concept such as the average length of life in a varied population which does not correspond to any particular group but is the arithmetic expression of the result; while the word 'Mean' should be used to denote some objective quantity such as the mean height of Englishmen about which all height measurements are grouped in a definite manner. The Arithmetic mean or average of the earnings of A in the last eight months which were Rs. 62, 72, 84, 93, 54, 64, 59, 64, is 69. In short, it is the quotient of the sum of all the items of the group divided by their number.

Computation of the Mean:—The computation of the mean is very simple. All that is to be done is to divide the sum of all the values of the items by their total number. This can be further simplified by using the short-cut method which consists in taking an arbitrary number and then tabulating the deviations of the actual numbers from such arbitrary number. The average of such deviations can be ascertained by dividing their total by the number of items and by adding it to the assumed number.

Computation of Mean by the *Short Cut Method*

Earnings Rs.	Arbitrary Average Rs.	Deviations from the arbitrary average
2784	2800	–16
2800	2800	0
2804	2800	4
2799	2800	–1
2794	2800	–6
2802	2800	2
2787	2800	–13
2814	2800	14
Total		–16

Average deviations $= -\frac{16}{8} = -2$

Thc mean is Rs. 2800 + (—2) = Rs. 2798.

If the arithmetic mean is to be computed from the frequency distribution table the sum of the items multiplied by the respective frequencies is to be divided by the total number of frequencies.

Computation of arithmetic mean

Heights in inches	Number of persons	Product
63	12	756
64	14	896
65	20	1,300
66	13	858
67	8	536
68	3	204
Total	70	4,550

$$\text{Arithmetic Mean} = \frac{4550}{70} = 65 \text{ inches.}$$

The Formula Method:—To minimise calculations in the case of a frequency distribution table it is convenient to use the following formula.

$$\text{Mean} = \text{Arbitrary Mean} + \frac{\Sigma fd}{n} \times i$$

where Σfd = the sum of the frequencies multiplied by their respective deviations from the arbitrary mean.

n = the total number of frequencies (f).

i = the class-interval.

In the case of classes it is difficult to know the exact values of the different variables in the same class and so as a matter of convenience mid-points of classes are taken as values of frequencies falling in that class. Again to further simplify the calculation work the deviations are taken by classes and not by units. The adjustment is made later on in the formula by multiplying it by the class–interval as shown in the above formula. This adjustment is known as 'Correction Factor.' It must be carefully noted that if there is no frequency for any class in the middle of the series while writing the deviations by classes the value of the deviations for a such a class must be assigned. For instance, if in the following example there would be no student who has obtained marks between 40 and 45 and because of that, if this class is not written the deviation for the previous class, i. e, 35–40 should be–3 and not–2. The easy way is to write all the groups and to write zero against it in the frequency column to avoid possible error.

Computation of Mean of marks of 100 students of Sec. 1.

Marks obtained	Mid-Point	No. of students f	Deviations d	fd
30–35	32.5	8	–4	–32
35–40	37.5	5	–3	–15
40–45	42.5	2	–2	–4
45–50	47.5	19	–1	–19
50–55	52.5	8	0	0
55–60	57.5	20	1	20
60–65	62.5	9	2	18
65–70	67.5	5	3	15
70–75	72.5	23	4	92
75–80	77.5	1	5	5
Total		100		80

In the above example 52.5 has been taken as the arbitrary mean and the deviations taken for classes are–1, –2, –3 and so on and not the *actual* deviation by units which would have been –5,–10, –15 and so on. By using the formula, we get,

$$
\begin{aligned}
\text{Mean} &= 52.5 + \frac{80}{100} \times 5 \\
&= 52.5 + 4 \\
&= 56.5
\end{aligned}
$$

It should be noted that the algebraic sum of the deviations of a series of magnitudes from their arithmetic mean is zero and that the arithmetic mean of two or more series can be computed from the following formula:

Mean of the combined series $= \frac{M_1n_1 + M_2n_2}{n_1 + n_2}$ where M_1 and M_2 are the arithmetic means of the series and n_1 and n_2 the number of frequencies.

Computation of the Mean of the marks of 90 students of Sec. II

Marks obtained	Mid-points	No. of students	Deviations	fd
30–35	32.5	10	–4	–40
35–40	37.5	8	–3	–24
40–45	42.5	18	–2	–36
45–50	47.5	3	–1	–3
50–55	52.5	4	0	0
55–60	57.5	19	1	19
60–65	62.5	18	2	36
65–70	67.5	3	3	9
70–75	72.5	5	4	20
75–80	77.5	2	5	10
Total		90		–9

$$\begin{aligned} \text{Mean} &= 52.5 + \frac{-9}{90} \times 5 \\ &= 52.5 - .5 \\ &= 52.0 \end{aligned}$$

The mean of the two combined series giving marks of students of section I and II can be ascertained as under by the formula given already.

$$\text{Mean} = \frac{(56.5 \times 100) + (52 \times 90)}{100 + 90} = \frac{5650+4680}{190}$$

$$= \frac{10330}{190} = 54.37$$

This result can be tested by computing the mean of the new series formed by combining the frequencies of the two existing series as shown below.

Computation of the Mean of marks of 190 students from sections I and II.

Marks obtained	Mid-point	No. of students	Deviations	fd
30–35	32.5	18	–5	–90
35–40	37.5	13	–4	–52
40–45	42.5	20	–3	–60
45–50	47.5	22	–2	–44
50–55	52.5	12	–1	–12
55–60	57.5	39	0	0
60–65	62.5	27	1	27
65–70	67.5	8	2	16
70–75	72.5	28	3	84
75–80	77.5	3	4	12
Total		190		–119

$$\text{Mean} = 57.5 + \frac{-119}{190} \times 5$$

$$= 57.5 - 3.13$$

$$= 54.37$$

Effect of grouping upon the arithmetic mean:–When the arithmetic mean is calculated each frequency is multiplied by the mid-point of its respective group and this is done under the assumption that the items in that group are equally distributed which is not invariably the case. It is because of this that the mean thus computed cannot be very accurate and again it is very likely that different class intervals will give different arithmetic means as the computation of the mean depends upon the method of grouping.

Computation of the mean of the marks of 190 students from section I and II by taking 10 as the class interval.

Marks obtained	Mid–point	No. of students	Deviations d	fd
30–40	35	31	–2	–62
40–50	45	42	–1	–42
50–60	55	51	0	0
60–70	65	35	1	35
70–80	75	31	2	62
Total		190		–7

$$\text{Mean} = 55 + \frac{-7}{190} \times 10$$
$$= 55 - 0{\cdot}368$$
$$= 54{\cdot}63$$

The arithmetic mean obtained by taking 10 as class interval is 54.63 while in the previous case it is 54.37. This difference in the result is due to taking different class intervals. In the previous example it is 5 while in this it is 10.

Unequal class intervals:—Sometimes arithmetic mean is required to be calculated from frequency tables with unequal class–intervals. In such a case the short–cut method of taking the deviations by classes cannot be used and they will have to be taken by units as illustrated in the following example.

Salary	Mid-point	Frequency f	Deviations from the Arbitrary means d	fd
40–50	45	30	−40	−1200
50–70	60	16	−25	−400
70–100	85	19	0	0
100–110	105	20	20	400
110–120	115	10	30	300
120–140	130	5	45	225
Total		100		−675

$$\text{Mean} = 85 + \frac{-675}{100} = 85 - 6.75 = 78.25$$

The question of unequal class interval becomes more perplexing in the case like the following:—

Salary	Frequency
Below 50	30
50– 70	16
70–100	19
100–110	20
110–120	10
120 and over	5

In this case the difficulty is regarding the mid-points of the first and last groups having a very wide range. Approximate limits can be placed based on the available information. Once this is done the mean can be computed by the method adopted in the previous case.

Quadratic Mean:—It is the mean obtained by extracting the square-root of the arithmetic mean of the squares of the items contained in the series. Thus the Quadratic mean of 6, 7 and 8 will be $\sqrt{\frac{6^2+7^2+8^2}{3}}$ = 7 05. It is also called Root-mean square.

Composite Averages:—Composite Average is computed by adding the simple averages and dividing the sum by the number of simple averages taken.

Monthly sales of the Modern House Ltd. for 1938.

	Rs.		Rs.		Rs.
Jan.	14320	May.	16360	Sept.	23990
Feb.	13908	June.	17470	Oct.	26580
March	18400	July.	24300	Nov.	19280
April.	19760	August.	22680	Dec.	20312

Total for the year Rs. 237360

$$\text{Monthly Average} = \frac{237360}{12} = \text{Rs. } 19780.$$

The monthly average for the year 1938 is Rs. 19780 and when such simple averages for few years are collected the composite average is computed as under:—

Monthly average sales for six years.

Year	Monthly average of Sales
1938	Rs. 19780
1939	20470
1940	21308
1941	19408
1942	20200
1943	22740
	Total Rs. 123906

Composite average = 123906 ÷ 6 = Rs. 20651

The Weighted Arithmetic Average:—So far each item has been counted once only and hence the average obtained is a Simple Average. Some items command greater importance and so in the computation of the arithmetic average they are assigned weights, i. e. they are counted more than once and the average so computed is known as *Weighted Arithmetic Average.*

The weightage applied may vary in different cases according to the relative importance. This distinction will be clearly seen from the following example:—

Salaries of teachers in towns A and B

Schools	Town A			Town B		
(1)	No. of teachers. (2)	Rate of Salary Rs. (3)	Product (2) × (3) Rs. (4)	No. of teachers. (5)	Rate of Salary. Rs. (6)	Product (5) × (6) Rs. (7)
Municipal School	15	35	525	35	50	1750
Government School	25	45	1125	35	58	2030
Aided School	20	33	660	10	21	210
Non-aided School	17	32	544	12	22	264
Night School	12	30	360	9	24	216
Total	89	175	3214	101	175	4470

In both the towns there are five types of schools and the simple and weighted arithmetic averages of salaries in towns A and B when computed will be as under.—

	Town A	Town B
Simple Arithmetic Average	$=\frac{175}{5}=$ Rs. 35	$\frac{175}{5}=$ Rs. 35
Weighted Aritmetic Average	$=\frac{3214}{89}=$ Rs. 36.1	$\frac{4470}{101}=$ Rs. 44.2

It will be clear from the above that though the simple average is the same in both the towns the weighted Arithmetic average is different. This type of average is often used when comparisons are to be made regarding the marriage, birth and death rates of different places or of different times.

Birth and Death Rates:—Birth and death rates are expressed as per thousand of the mid-year population of the place and are usually given upto one place of decimals. But these are *the General or Crude Rates* and for making accurate comparisons with figures of different places a method of correction or standard must be applied. This is particularly necessary in the case of death rates as the average mortality at different ages has a great influence on the crude rates and so this problem can be better solved by arriving at corrected rates by the application of weights.

Corrected or Standardised Death Rates:—When the death rates at different places are to be compared the crude or general death-rates fail to give a correct picture of the situation as the influence of the average mortality at different ages is not given proper weightage. This defect is removed by ascertaining a corrected or standardised death–rate as this is done under the assumption that the distribution of population in different age-groups in any place is not abnormal but is the same as that of the other town with which its death-rate is being compared. If the death rates of the two towns, A and B, are to be compared the death-rate of one town will be corrected to make an effective comparison. When we want to have the corrected Death Rate of town A for comparison with that of the town B we shall multiply the death rate of each age-group of the town A by the corresponding number of people in town B, in the same group. The total will then be divided by the total population of the town B, thus obtaining the corrected death rate of the town A as compared to the death rate of the town B as will be seen from the following illustration :—

	Town A			Town B		
Age group	Population	Deaths	Death-rate	Population	Deaths	Death rate
Under 5	25,000	550	22	10,000	220	22
5–15	40,000	280	7	15,000	105	7
15–35	60,000	720	12	20,000	240	12
Over 35	15,000	525	35	15,000	525	35
Total	1,40,000	2075	14.8	60,000	1090	18.1

General or Crude Death Rates (weighted average)

$$\text{Town A} = \frac{(25000 \times 22) + (40{,}000 \times 7) + (60000 \times 12) + (15000 \times 35)}{25000 + 40000 + 60000 + 15000}$$

$$= \frac{2075000}{14000} = 14{\cdot}8$$

$$\text{Town B} = \frac{(10000 \times 22) + (15000 \times 7) + (20000 \times 12) + (150000 \times 35)}{10000 + 15000 + 20000 + 15000}$$

$$= \frac{0190000}{60000} = 18{\cdot}1$$

From the study of the general death rates it is seen that the town A is better, but when the corrected death rate of the town A as compared to that of B is ascertained it is seen that this is not the case.

Corrected Death Rate of town A as compared to that of B

$$= \frac{(22 \times 10000) + (7 \times 15000) + (12 \times 20000) + (35 \times 15000)}{60000}$$

$$= \frac{1090000}{60000} = 18{\cdot}1$$

From this it will be seen that both the towns are equally healthy.

This result can be verified by finding the corrected death rate of town B as compared to that of A.

Corrected Death Rate of town B as compared to that of A

$$= \frac{(22 \times 25000) + (7 \times 40000) + (12 \times 60000) + (35 \times 15000)}{1{,}40{,}000}$$

$$= \frac{2075000}{140000} = 14{\cdot}8.$$

From this it will be seen that for making proper comparisons we should not take the general death rate but the corrected one.

The Descriptive Average:–It is a type of an average which is not quite reliable but is compiled more to serve as a guide because the data available are either inexact or incomplete. Supposing the weights of 200 persons are given as 10 having weights between 120 and 130 lbs., 25 between 130 and 140 lbs. 30 between 140 and 150 lbs., it becomes necessary to assume that their weights in the particular group are equal which is rarely the case. By taking the average weight of 10 persons as 125 we may calculate the combined weight of the group as 1250 lbs. which is not likely to be absolutely correct. Such an average is known as 'Descriptive Average' and though not quite reliable can be used with advantage in cases of equally indefinite nature.

Advantages and Disadvantages:—Of all the types of averages the arithmetic mean is more commonly used and is even known as 'Common Average' or simply 'Average.' It is more known to an average man as compared to the median or the mode and is also rigidly defined. It is very easy to compute and, unlike the median or the mode, not only takes all the items into account but gives proper importance to the items on the extremes. It can be used for further mathematical calculations. If, for instance, the average income of each person as well as the total number of persons is known, the total income can be easily ascertained. This is not possible in the case of other form of averages such as the median or the mode.

The disadvantages of the arithmetic mean are that it does not usually correspond to any observation in the series and as such it is called a *fictitious* average. If the daily average of the number of persons employed in a mill is calculated it often includes some decimal figure but how can there can be fraction of a man at work? Again if the correct value of even a single item in the series is not available it is not possible to compute the mean. It gives sometimes too much importance to stray items on the extremes. In the wages bill of an establishment because of the high salary of the head the value of the mean will go up, thus giving undue importance to the salary of one individual. If the samples are not properly selected the mean computed therefrom is sure to fail to give a true picture of the series. It can neither be located on a frequency graph, as is done in the case of the median or the mode, nor obtained by mere inspection. If the actual facts are not known in detail the mere knowledge of the arithmetic mean is likely to be deceptive and in some cases dangerous as will happen to the person who would try to cross the river knowing the average depth of water to be three feet and failing to realise that at one place the depth is seven feet!

The Median

When the observations are arrayed *i. e.* arranged in order of their size the median is the value of that item which has equal number of observations on either side. If we are required to ascertain the median of the weights of eleven persons all the items will have to be first arrayed in order of their values as under :—

132, 137, 139, 143, 145, **147,** 152. 160, 163, 167, 175

As there are eleven items the median number will be the sixth as in that case there are five items above and five below it. The value of the sixth item is 147 and so the median weight obtained is 147 lbs. It is visible from this that the median has two equal divisions, one has all values greater while the other has all values smaller than the median. In this case the number is odd and so it is easy to make two equal divisions but in the case of an even number of items the median is intermediate between the sizes of the two middle items. In such a case the average value of these two items is taken as the value of the median.

Computation of the Median:—In case all the recorded observations are given they can be arranged in ascending or descending order and the value of the middle item ascertained is then the median. But as we are usually supplied with the items in the form of a frequency table the median is computed by using one of the following methods:—

(i) The Formula Method.
(ii) The Graphic Method.

The Formula Method:—In the case of the data given in a frequency distribution table the median cannot be ascertained by arraying the items as can be done in the case of the original data. It can, however, be computed with reasonable accuracy by interpolation and for this the formula to be used is as under :—

$$\text{Median} = 1 + \left(\frac{\frac{n+1}{2} - cf}{fr} \right) \times i$$

Where 1 = the lower limit of the class in which the median item is situated.

n = the total number of items.

cf = the cumulative frequency of the group lower to the one of the median number.

fr = the frequency of the class of the median number.

The median item or the median number can be ascertained by using the formula $\frac{n+1}{2}$.

Computation of Median

Wages	No. of workers or frequency	Cumulative frequency
Rs. 25 and under Rs. 30	10	10
30 – – 35	15	25
35 – – 40	19	44
40 – – 45	26	70
45 – – 50	18	88
50 – – 55	29	117
55 – – 60	24	141
60 – – 65	4	145
65 – – 70	8	153
70 – – 75	7	160

The median number is $\frac{160+1}{2}$ i. e. 80·5 and it is situated in the class 45–50. By using the given formula, the median will be computed as under :—

$$\text{Median} = 1 + \left(\frac{\frac{n+1}{2} - cf}{fr} \right) \times i$$

$$= 45 + \frac{80{\cdot}5 - 70}{18} \times 5$$

$$= 45 + 2{\cdot}91$$

$$= 47{\cdot}9$$

The Graphic Method:—After constructing a cumulative frequency curve a horizontal line paralled to the base should be drawn from the mid-point of the ordinate representing the total frequency of the series. From the point where this line cuts the cumulative frequency curve a perpendicular to the base should be drawn. The median is the distance between the origin and the foot of the perpendicular. The other method is of drawing two cumulative frequency curves one based on 'More than, basis and the other on 'Less than' basis and from the point of their intersection a perpendicular should be drawn to the base line. The median then will be obtained by taking the reading from the foot of this perpendicular as in the previous case. (vide Fig. 3 in the chapter on Graphs).

The median can also be ascertained by drawing a simple frequency curve. In this case a perpendicular is required to be drawn to the base bisecting the total area covered by the frequency curve and the reading on the base line at the foot of the perpendicular is the Median. This method, however, is difficult and hence is not generally used.

It can also be located by Galton's method by marking observations on a piece of graph paper by dots as explained in the chapter by Graphs (Fig. 9). This method, however, is not quite satisfactory.

Advantages and disadvantages:–The advantages claimed for the median as a type of average are many. Besides being very easy to calculate it is neither affected by the items on the extremes nor is it necessary to have details about them for computing it. It is difficult to calculate the arithmetic mean if the class intervals are unequal but the median can be easily computed in case of such data. In the case of a qualitative data the median can be ascertained by arranging the variables according to the quality which is not possible in the case of other forms of averages. From the point of view of intelligence the median student can be ascertained by arraying the group of students based on their quality of intelligence but this cannot be done for ascertaining the mean. Lastly, a visual representation is possible as the median is usually an actual example while the mean is often visionary. The main defects of the median are that it is not possible to obtain the actual median in the case of a group having an even number of observations and so in such a case as a makeshift an average of the two items in the middle is taken. Again very little importance is attached to items on the extremes and as such the median fails to register charges due to the changes in the values of the items on the extremes. It gives no basis for further calculation. Even if the salary of a median worker is known it is not possible to find the total salary bill when the number of workers is given. In the case of mean, this can be easily calculated. It is likely that the median may be located at a point in the series where there are few or no observations, and thus it may fail to represent the real state of affairs. In general, however, it is a very useful form of an average particularly for the study of wages and that of distribution of wealth.

Quartiles

The median is the value of the item at the centre that divides the arrayed observations into two equal parts while in the case of quartiles four equal divisions are made. If the observations are arrayed in the ascending order of magnitude the item having a position midway between the lower extreme and the median is the Lower or First Quartile and that between the median and the upper extreme is the Upper or Third Quartile. The median is thus one of the quartiles and is often known as Second Quartile.

These quartiles like the median can be ascertained by the formula or by the graphic method, the only difference being that the median is a midway point while these are the midway points between the median and the two extremes. When the series is arrayed in an ascending order the quartiles are computed by the following formula :—

$$\text{Lower Quartile i. e. } Q_1 = 1 + \frac{\left(\frac{n+1}{4}\right) - cf}{fr} \times i$$

$$\text{Upper Quartile i. e. } Q_3 = 1 + \frac{\frac{3(n+1)}{4} - cf}{fr} \times i$$

But in the above formulae 'cf' is the cumulative frequency of the group lower to the one in which the particular quartile is situated and 'fr' is the frequency of the class in which the particular quartile is situated.

In the foregoing frequency table giving wages of 160 workers the lower and upper quartiles are 39.0 and 55.78 respectively as will be seen from the solution given below :—

The lower quartile number is $\frac{160+1}{4}$ i.e. 40.25 and it is situated in the class 35–40 while the Upper quartile number is $\frac{3(160+1)}{4}$ i.e. 120.75 and is in the class 55–60.

$$Q_1 = 35 + \frac{\frac{(160+1)}{4} - 25}{19} \times 5$$

$$= 35 + \frac{40.25 - 25}{19} \times 5$$

$$= 35 + 4.01$$

$$= 39.0$$

$$Q_3 = 55 + \frac{\frac{3(160+1)}{4} - 117}{24} \times 5$$

$$= 55 + \frac{120.75 - 117}{24} \times 5$$

$$= 55 + 0.78$$

$$= 55.78$$

These quartiles can be ascertained by the graphic method by drawing parallel lines to the base from the points in the ordinate that are midway between the median and the two extremes. From the two places where the cumulative frequency curve is cut by these parallel lines perpendiculars are to be dropped to the base and from their meeting points to the base line the two quartiles can be ascertained. (See Fig. 3 in the chapter on Graphs).

Quintiles, Octiles, Deciles and Percentiles

The Quintiles, Octiles, Deciles and Percentiles are other forms of minor importance and are to be calculated by the method used for computing the quartiles with the changes necessary to suit the particular form. The quintiles divide the group into five equal parts while the octiles, deciles and percentiles divide it into eighths tenths and hundredths respectively. These values of the variables can also be ascertained by the graphic method just as is done in the case of quartiles.

The Mode

It is usually found that in a given data a certain magnitude will occur more frequently than any other and this predominant group can be easily located. The value of the item which is so common is known as Mode or 'Norm'. It automatically follows that if items are selected at random and not deliberately from the data the most likely item to occur will be of model value. Even in the ordinary use mode means prevailing fashion or style. The standard of passing the matriculation examination being 33 percent the greatest number of students will usually figure in the class of those getting between 30 and 40 percent marks which will then be the model class in the frequency table showing the distribution of students according to marks obtained at the matriculation examination. The Mode, in short, is the size of the variable which occurs most frequently. It is the point of great density in the frequency table. In the sample observations taken, the *apparent mode* is the model class and from that we arrive at the *computed mode* by certain calculations while the *true mode* may be still different. Sometimes there are more than one mode if there happen to be more classes with the same frequency and in such a case if frequency distribution curve in drawn it will not be *Unimodal* but *multimodal.*

Computation of the Mode:—If all the recorded observations are available before they are tabulated the mode is the size of that item which has occurred for the highest number of times. However, we are usually supplied with the observations in a tabulated form and so the mode or the *Inspectional Average,* as it is sometimes callled, can be ascertained by one of the following methods :—

(i) The Group Method.
(ii) The Formula Method.
(iii) The Graphic Method.

1. **The Group Method:**—The Bracketted Method or the process of Grouping thongh not very accurate is particularly useful when the mode is not clearly present or when the series happen to possess more than one mode. In this method the procedure consists in grouping the frequencies normally in twos and threes. In the beginning the items grouped are first and second, third and fourth and so on. In the second case they are grouped by leaving the first, and then by taking them as second and third, fourth and fifth and so on. Next the grouping made is of three items such as, first, second and third and then fourth, fifth, and sixth. The first item is left in the next attempt and the group taken is of second, third and fourth and so on. In each case the classes having the highest number in the group are noted and the mode is then said to lie in the class which is found to have formed the group of classes securing the highest number. The grouping done is normally by twos and threes but if necessary the grouping can be done by taking four or more items.

Location of Mode by the Process of Grouping.

Height in inches.	f	Frequency				
		In groups of two		In groups of three		
	1	2	3	4	5	6
60–61	14	30		38		
61–62	16		24		48	
62–63	8	32				**67**
63–64	24		**59**	**77**		
64–65 ✓	**35**	**53**			**72**	
65–66	18		37			53
66–67	19	35		45		
67–68	16		26			
68–69	10					

If the results of the above grouping are tabulated they will be as under :—

Column	Groups having the maximum frequencies				
1			64—65		
2			64—65	65—66	
3		63—64	64—65		
4		63—64	64—65	65—66	
5			64—65	65—66	66—67
6	62—63	63—64	64—65		
Number of times	1	3	6	3	1

It is evident, that the the group 64—65 is the most frequent and is, therefore, the Modal group.

This method is more applicable in the case of the data having a discrete series but in a continuous series it will give only the class in which the mode is located.

2. **The Formula Method:**—This formula method of determining the Mode is based on the assumption that the position of the mode in the model class is determined by the relative sizes of the adjacent groups. This is emperical method and the formula used is as under:—

$$\text{Mode} = 1 + \frac{f_2}{f_2 + f_1} \times i$$

where 1 is the lower limit of the class in which the apparent mode is situated.

f_1 is the number of items in the class below that containing the mode.

f_2 is the mode number of items in the class above that containing the mode, and

i is the class interval.

By using the above formula to find the mode from the table giving the heights of 150 persons we get,

$$\begin{aligned}\text{Mode} &= 64 + \frac{18}{18+24} \times i \\ &= 64 + 0.43 \\ &= 64.43\end{aligned}$$

By the above formula the mode has been calculated by taking the lower limit of the model class and by working upwards. We can, however, get the same result by working downwards but in that case the formula used will have to be as under and 1 will be the upper limit of the modal class.

$$\text{Mode} = 1 - \frac{f_1}{f_1 + f_2} \times i$$

By usidg this formula we will have

$$\text{Mode} = 65 - \frac{24}{24 + 18} \times i$$
$$= 65 - 0.57$$
$$= 64.63$$

There is yet another better formula based on the relationship between the values of mean median and mode but for this the values of mean, and median will have to be computed beforehand.

Mode = Mean − 3 (Mean − Median).
To be more accurate 3 should be replaced by 3·03

The Graphic Method: -- If a frequency graph is constructed from the frequency table the abscissa of the highest point of the curve will give the modal size as will be seen from Fig. 2 in the chapter on Graphs. If a cumulative frequency graph is constructed the abscissa of the point where the curve is steepest will represent the modal size.

Uses of Mode: -The utility of mode is being appreciated more and more day by day. In the science of Biology and Meteorology it has been found to be of great value. Of course, in commerce and industry too it is attaining a very great importance. If the data is carefully and properly collected the modal item can be ascertained with a fair accuracy and it gives a good basis for calculations and comparisons. The work of a modal man or machine can be taken as the basis for estimating the output and deficiencies can thus be ascertained. If a modal foot can be measured the employees in a shoe factory, when they have not got orders on hand, can produce pairs of shoes to fit it. This will help to keep the labour off from unemployment in the slack season. Many establishments are these days engaging their attention in keeping statistics of their sale to ascertain the particulars of the modal articles sold.

Advantages and disadvantages: --As a form of an average the mode is very useful. Not only it gives a good picture of the type of distribution but it can be easily and clearly understood by an average person. Very often it can be ascertained by mere inspection and is known as an '*Inspectional average*'. Even if it is computed by a formula the labour and time required is much less as compared to that required in the case of the arithmetic mean. It is neither necessary to know the details of the extreme items nor do they affect the mode unless they are in the modal class which is very rarely the case. It can in fact be compared by making use of few items only. Unless the series is a continuous one it represents an actual case as is done by the median but it has over the median the advantage of being of the predominating magnitude. Besides, it is found to be of great utility in scientific and commercial matters.

But the mode is defective in some respects. In the first place there are very often more modes than one in the series and in such a case not only its computation is difficult but even its utility is practically nominal. In an irregular group it loses its value. It fails to give proper importance to extreme items and as it is calculated by taking only a few items into consideration it is not based on all observations as is done in the case of mean. It does not tend to further arithmetical treatment. Even if the mode is known the aggregate of the series cannot be obtained in spite of knowing the total number of items. In fact as a clearly defined mode does not always exist and so its utility is not of much importance.

The Geometric Mean

The Geometric Mean or Average is the root of the product of all the observations comprising the group and it is computed by the following formula.

$$\text{Geometric Mean} = \sqrt[n]{a \times b \times c \times d \times e}$$

where n is the total number and a, b, c, d are different items. Thus the Geometric Mean of numbers 1, 2, and 4 is

$$Mg = \sqrt[3]{1 \times 2 \times 4} = 2$$

If however there is weightage present in the series of n observations the following formula will have to be used.

$$Mg = \sqrt[n]{a^{a_1} \times b^{a_2} \times c^{a_3} \times d^{a_4}}$$

where a, b, c, d will be the items and a_1, a_2, a_3 a_4 the corresponding weights.

Computation of Geometric Mean

Class interval	Midvalue	Frequency
25–35	30	2
35–45	40	3
45–55	50	4
55–65	60	1
Total		10

$$\text{G. M.} = \sqrt[10]{30^2 \times 40^3 \times 50^4 \times 60^1}$$

For the computation of the geometric mean it is necessary to use logarithms to avoid calculations. In the first case i.e. in the case of series without weightage the formula will be—

$$\text{Log Mg} = \frac{\log a + \log b + \log c}{n}$$

$$\text{i.e. Log Mg} = \frac{\log_1 + \log_2 + \log_4}{3}$$

$$= \frac{.00 + .3010 + .6021}{3}$$

$$= \frac{0.9031}{3} = 0.3010$$

$$\text{Mg} = 2.0$$

In short the arithmetic mean of the logarithms of the individual measures will give the logarithm of the geometric mean.

In the case of weightage the formula will be

$$\text{Log Mg} = \frac{a_1 \log a + a_2 \log b + a_4 \log c}{n}$$

Using this formula in the case of the above example we get

$$\text{Log Mag} = \frac{2 \log 30 + 3 \log 40 + 4 \log 50 + 1 \log 60}{10}$$

$$= \frac{2\ 9542 + 4.8063 + 6.7960 + 1{,}7782}{10}$$

$$= \frac{16.3347}{10} = 1.63347$$

$$\text{Mg.} = 43\cdot0$$

The Geometric mean is much less than the arithmetic mean unless all the quantities are equal which is rarely the case. The advantages claimed for this form of average are that it takes all observations into account, the weight attached to large items is less as compared with the arithmetic mean and it tends to further arithmetic and algebraic manipulations. But the great drawback is that it cannot be computed if any of the quantities is zero as the whole result will then be zero. It will not generally point out any observation in the series and for an average man it will be difficult to comprehend it apart from the fact that it is very difficult to compute. Again it gives comparatively less weight to extreme variations.

Jevons used it for his study of prices. It is also being used for the Board of Trade Index Number of Wholesale Prices by the Board of Trade. In India it is made use of by the Calcutta Weekly 'Capital' for computing the 'Index of Business Activity in India.' But on the whole it is not much used.

The Harmonic Mean

The Harmonic mean is computed by calculating the reciprocals of the arithmetic average of the reciprocals of the items of the series and the formula to be used is

$$\text{Mh} = \frac{n}{\frac{1}{a} + \frac{1}{b} + \frac{1}{c}}$$

where n is the total number of items and a,b,c...the items in the series.

The harmonic mean of the following four items 3, 5, 9 and 10 will be

$$Mh = \frac{4}{\frac{1}{3} + \frac{1}{5} + \frac{1}{9} + \frac{1}{10}} = \frac{4}{\frac{67}{90}} = 5.37$$

The harmonic mean which is smaller than the geometric mean is useful in measuring the change or the variation in the data and in its computation extreme items are not given much weight. It is however, difficult to comphrehend its meaning and also to calculate because of the calculations involved. It has a restricted field and is not made much use of by statisticians.

Relations between different forms of averages:—When all the averages are ascertained for a given series of magnitudes it will be seen that there exists certain relationship amongst them. In the case of a perfectly symmetrical distribution the mean, the median and the mode will coincide. In a moderately symmetrical distribution the position of the median is between the mean and the mode occupying the place one-third the distance along the scale from the mean towards the mode. The geometric mean is lower in value than that of the arithmetic average but more than that of the harmonic mean, unless all the measures in the series are equal. The geometric mean of any two terms is equal to the geometric mean of the harmonic and the arithmetic means of these terms.

Uses of Averages:—From the description of the various forms of averages it will be seen how useful they are in statistics. Most of them play such an important part that some writers have styled though not quite correctly statistics as the science of averages. With the increasing use of statistics not only in scientific matters but in other spheres of life averages have been found to be of great utility in various ways.

In the first place they are very useful in giving a concise picture of a large group. Instead of looking to a large number of tables giving the rainfall for various years a fair idea of the rainfall of that place can be obtained from its average rainfall. It is very difficult to form an idea of the level of the class of 150 students by going through the sheets giving their results. If the percentage of passes in different subjects or the average marks obtained are given immediately a fair idea regarding the class is obtained. If the average wage of a labourer is given one can have a picture of the large group of labourers. *Secondly*, experience has shown that if an average is taken of the observations selected at random it comes almost close to the average of the whole series. Difference there may be but it is almost nominal. In short, with the help of averages one can manage with samples rather than go in for detailed study of the whole group. *Thirdly*, the averages make the work of comparing the two groups much easier. If the average wages of the labourers in two mills are ascertained it is sure to show which mill pays better wages. With the help of averages comparison of the past and present can be made easily and they can even form the basis for future comparisons. *Fourthly*, by the use of averages the influence of the extreme items is to a great extent reduced. In a mill the office boy may get Rs. 20/- per month while the manager may be drawing a fat salary of rupees two thousand. If averages can be made use of the influence of both these items on the extremes is greately minimised. *Lastly*, averages serve useful purpose of giving us the means to ascertain whether the items in the given group are substantialy uniform in size or there are great variations. This is achieved with the help of the methods of dispersion that are mainly based on averages. Again the averages tend to reduce the temporary irregular fluctuations as will be seen from the study of the moving averages. In the end a mathematical concept to the relationship between the different groups is given by the averages. The wages in a certain mill may be on the whole more than in the other but the computation of an average is the only method to give us a definite ratio of one to the other.

No doubt the averages are very useful but they have certain defects too. In the first place the averages give a fair idea of the situation but it is sometimes very misleading. They fail absolutely to give an idea of the trend. It is possible that two companies may have the same average of profits for the last five years yet one may be flourishing while the other may be on the verge of bankruptcy. Similarly, the average rainfall at two places during the season may be the same yet in one the rainfall might have been properly spread over and in the other the average might have been made up because of heavy rains on some days in the beginning and at the end of the saason. The condition of the crops at the two places estimated on their average rainfall will be erroneous. Lastly, most of the averages rarely correspond to any item in the series and hence they are more imaginary than real. However, if used with proper care they are very useful in scientific and commercial activities, particularly when used with the help of dispersion and skewness.

Essential requisites of an ideal average:—Averages are computed with a view to presenting a concise picture of the mass of the matter collected and hence it is necessary that an average should possess the characteristics essential for serving this purpose adequately. *Firstly,* the average to be computed should be based on the whole data as in the absence of this property it will fail to register the characteristics of the whole distribution. In the case of the median and the mode the items on the extremes absolutely fail to influence the average computed. *Secondly,* it should be rigidly defined and there should be definiteness about its calculation. It should in no way depend upon individual judgment as very often judgments differ in case of different persons depending upon their mood and temperament. In the absence of this there remains a scope for misconstruction and great difficulty is experienced in comparing them. *Thirdly,* the average, howsoever useful it may be, should be such as can be computed with reasonable ease and rapidity. In spite of the geometric mean being a better form of average it is not so popular as it is not easy to calculate. However, it is not proper to subordinate the utility of the average to the labour saving motive. *Fourthly,* it should be based on the large range of data with a view to minimising the influence of abnormal observations. It is of course natural that calculations based on different observations will give different values for the averages but the form of average should be such as will not yield greater difference in such cases. In short, it should be capable of not being much influenced by fluctuations in sampling and hence should be stable.

in character. *Fifthly*, it should have such simple properties as are likely to be understood easily by an average man and it should not be too absract like the geometric or the harmonic mean. *Lastly*, it should admit cf further mathematical treatment and not like the mode and the median which miserably fail to satisfy this desideratum.

If the above tests are applied in the case of the various averages it will be seen that the arithmetic mean is based on all observations, is rigidly defined, can be computed with greater case, is not likely to be much affected by fluctuations in sampling, is very easily understood and tends itself to further mathematical treatment. The median and the mode are well defined, can be calculated with great ease and can also be comprehended with fair ease but they do not give any importance to the extreme items, are likely to be affected by fluctuations in sampling and do not tend themselves to further mathematical calculations. The remaining two forms of averages i. e. the geometric mean and the harmonic mean are based on all observations, are rigidly defined and can lend to further mathematical treatment but they fail to be comprehended by an average man and the calculations involved in their computation prove a great handicap. Judging from the various standpoints the arithmetic mean is on the whole the best form of average, is also most popular and is rightly known as the 'Common Average.'

Exercises.

(*a*) MEAN

1. What is an average. ? Mention the important forms of averages and state which of them are most used.
2. Differentiate between (a) simple and weighted averages, and (b) composite and descriptive averages.
3. Discuss the merits and demerits of arithmetic mean as a form of average.
4. From the following table calculate the average price of a lb. of tea and also the weighted average.

Price per lb.	lb. sold.
0–14—0	96
1—0—0	48
1—4—0	36
1—7—0	20
1–10—0	15

5. (*a*) What is a weighted average? For what purpose is it used?

(*b*) From the following data relating to paper consumed by a press, find out the difference in the weighted average cost of paper for the two years:

Description of Paper	1942–43 Rate per lb.	1942–43 Quantity consumed	1943–44 Rate per lb.	1943–44 Quantity consumed
	Rs. a. p.	Tons	Rs. a. p.	Tons
White	0 – 7 – 2	17	0 – 8 – 6	11½
Badami	0 – 6 – 6	6	0 – 7 – 6	8½
Other kind	0 – 13 – 0	14	0 – 15 – 0	10
		37		30

(1946–Oct..)

6. The following are the marks of 93 students examined in History. Tabulate them by giving a class–interval of 10 and then find the arithmetic mean.

74,	14,	29,	36,	15,	49,	54,	69
77,	96,	14,	27,	3,	81,	17,	21
34,	72,	4,	22,	19,	92,	25,	34
67,	18,	57,	52,	28,	75,	2,	12
87,	35,	47,	61,	60,	43,	89,	91
27,	42,	47,	59,	68,	5,	33,	22
84,	97,	78,	59,	46,	34,	72,	65
47,	58,	63,	71,	75,	82,	17,	4
14,	92,	88,	77,	68,	54,	15,	29
17,	1,	9,	14,	15,	55,	67,	95
49,	27,	29,	48,	59,	64,	75,	87
19,	1,	97,	14,	54.			

7. From the following table find out the average earnings of the mill labourers in the Sabarmati mills.

Monthly wages. Rs.	No. of labourers.	Monthly wages. Rs.	No. of labourers.
15—18	21	39—42	36
18—21	29	42—45	45
21—24	19	45—48	27
24—27	39	48—51	48
27—30	43	51—54	21
30—33	94	54—57	12
33—36	73	57—60	5
36—39	68		

8. The weights of the number of the Bombay Gymnasium are shown in the following table. Find the Arithmetic Average by taking 137·5 lb. as the Arbitrary Mean.

Weights in lb.	Frequency	Weight in lb.	Frequency.
Below 100	5	135—140	27
100—105	31	140—145	42
105—110	29	150—155	29
110—115	93	155—160	13
120—125	105	160—165	3
125—130	72	170—175	4
		175—180	1

9. Heights of 310 students of the Albert College are given below and you are asked to find out the average height by the short-cut method and verify the result by the other method.

Height in inches	No. of Students	Height in inches	No. of Students
Below 57	12	63—64	42
57—58	19	64—65	12
58—59	29	65—66	17
60—61	47	66—67	5
61—62	92	68—69	1
62—63	34		

10. The attendance of outdoor patients for 60 days in the Rajabai Dispensary is as under. Find out the average attendance by taking 5 as the class interval.

105, 92, 87, 115, 124, 130, 95
147, 117, 95, 122, 130, 77, 87
108, 114, 141, 139, 96, 89, 104
89, 116, 112, 129, 114, 128, 123
137, 134, 93, 105, 103, 36, 101
100, 130, 112, 129, 134, 143, 149
91, 141, 107, 132, 113, 121, 120
107, 113, 115, 92, 131, 134, 108
112, 113, 99, 138.

11. From the following table find the average size of a field.

Area of fields in acres.	No. of fields.	Area of fields in acres.	No. of fields.
Below 1	87	8—8	12
1—2	55	8—9	10
2—3	43	9—10	8
3—4	37	10—11	12
4—5	28	11—12	4
5—6	19	12—13	7
6—7	17	14—15	1

12. The sale prices and the quantities of food grains sold in two towns are shown in the following table. Compare the weighted average cost per maund of food grains in the two towns :

Food grains	Town A		Town B	
	Retail price per maund Rs.	Qnantity sold per week in maunds	Retail price per maund Rs.	Quantity sold per week in maunds
Rice	14–8–0	50,000	20–8–0	11,600
Wheat	12–8–0	67,900	15–8–0	28,100
Jowar	7–0–0	5,600	—	—
Bajri	7–12–0	20,800	11–8–0	4,200

(1945–Oct.)

(*b*) MEDIAN.

13. Define a median and give the different methods of computing it.

14. Write short notes on :—
Quartiles, deciles and percentiles.

15. Discuss the comparative advantages and disadvantages of mean or median.

16. Calculate the median from the following data :—

Monthly pay in rupees.	No. of receipients.	Monthly pay in rupees.	No. of recipients.
Below 50	57	100—110	24
50—60	39	110—120	14
60—70	48	120—130	11
70—80	26	130—140	6
80—90	21	140—150	2
90–100	19		

17. The following table gives the number of deaths in different age-groups. Find the median and the upper quartile.

Age groups.	No. of deaths in a month.	Age groups.	No. of deaths in a month
Below 5	2		
5—10	126	35—40	25
10—15	47	40—45	14
15—20	29	50—55	19
20—25	57	55—60	28
25—30	65	60—65	12
30—35	32	65—70	11

18. From the information given regarding the population of 87 villages find the population of the median village.

Population.	No. of villages	Population.	No of villages
Below 500	27	2500–3000	5
500–1000	11	3000–3500	6
1000–1500	9	3500–4000	7
1500–2000	10	4000–4500	5
2000–2500	7		

Find the first decile also

19. From the following table giving the annual incomes of 3000 landlords find the median and the first quartile.

Income in rupees.	No. of landlords	Income in rupees	No. of landlords.
Below 1000	724	7000—8000	115
1000—2000	439	8000—9000	87
2000—3000	382	9000—10000	97
3000—4000	357	10000—11000	64
4000—5000	298	11000—12000	15
5000—6000	202	12000—13000	17
6000—7000	193	13000—14000	10

20. Determine graphically the quartiles and the median of the distribution given below :—

No. of paylees sold.	No. of ration shops.
200—299	5
300—399	9
400—499	15
500—599	17
600—699	22
700—799	17
800—899	7
900—999	6
1000—1099	2
Total	100

(1945–Mar.)

21. Find graphically the median and the third quartile of the following distribution;—

Weight in grams.	No. of apples.
60— 79	5
80— 99	13
100—119	16
120—139	8
140—159	4
160—179	–
180—199	2
200—219	1
220—239	1
Total	50

(1947–Oct.)

22. Determine graphically the quartiles and the median for the following table :—

Income.	No. of persons.
Below Rs. 30	69
Rs. 30 and below Rs. 40	167
Rs. 40 „ „ 50	207
Rs. 50 „ „ 60	65
Rs. 60 „ „ 70	58
Rs. 70 „ „ 80	27
Rs 80 and over	10
Total	603

(1942–Mar.)

Discuss the advantages and dis-advantages of the median as average.

(c) Mode.

23. Define mode and discuss its utility particularly in Commerce.

24. Discuss the different kinds of averages used in statistical methods and give their relative merits with illustrations.

(1932–Mar.)

25. Write short notes on :—

Unimodal curve, apparent mode and bracketted method.

26. Give the different methods of ascertaining the mode and state with reasons the method you would like to follow.

27. Define and discuss the properties and uses of the following forms of averages :—

Arithmetic mean, median, mode and geometric mean. (1943-Oct.)

28. The test-marks of 150 students are given below. Find the Mode.

Marks.	No. of recipients.	Marks.	No. of recipients.
Below 5	14	25—30	18
5—10	11	30—35	18
10—15	12	35—40	14
15—20	19	40—45	4
20—25	38	45—50	2

29. The earnings of 72 buses plying in Ahmedabad on 5th January, 1948 are as under :—

Rs.	72,	85,	93,	84,	105,	121,	71
	69,	92,	106,	68,	107,	112,	96
	104,	103,	87,	69,	85,	104,	92
	103,	92,	101,	89,	74,	69,	99
	104,	86,	103,	102,	97,	94,	91
	76,	82,	104,	102,	131,	117,	112
	128,	124,	103,	101,	102,	104,	94
	117,	112,	104,	102,	109,	110,	93
	87,	74,	84,	102,	101,	112,	107
	79,	81,	104,	103,	94,	97,	81
	101,	104					

Find the Mode by taking 5 as the class interval.

30. The record of sale of tea of the Capital Stores Ltd. is given below and you are asked to ascertain the mode.

Price per lb. in rupees.	No. of lb. sold.	Price per lb. in rupees.	No. of lb. sold.
Below 0–12	74	1—2 to 1—3	68
0–12 to 0–13	38	1—3 to 1—4	47
0–13 to 0–14	39	1—4 to 1—5	35
0–14 to 0–15	42	1—5 to 1—6	22
0–15 to 1—0	57	1—6 to 1—7	12
1—0 to 1—1	87	1—7 to 1—8	7
1—1 to 1—2	74		

31. From the following data find the Mode.

Height in inches	Frequency	Height in inches.	Frequency
Below 58	17	63—64	37
58—59	27	64—65	14
59—60	14	65—66	8
60—61	19	66—67	12
61—62	22	67—68	5
62—63	43		

32. Recast the following cumulative table into the form of an ordinary frequency distribution and determine the value of the mode by using the formula :

Mean–Mode=3 (Mean–Median)

No. of days absent	No. of Students	
Less than 5	29	
„ „ 10	224	
„ „ 15	465	
„ „ 20	582	
„ „ 25	634	
„ „ 30	644	
„ „ 35	650	
„ „ 40	653	
„ „ 45	655	(1946–Mar.)

33. Discuss the essential requisites of an average and in light of them critically examine all the forms of averages.

34. Wages of six labourers are given below and you are asked to calculate geometric and harmonic means.

Rs. 40, 50, 60, 70, 80, 90.

35 'The use of weights in the computation of weighted averages, which is a laborious process and continually leads to controversies, appears to be superfluous'. Critically examine this statement by giving suitable examples.

36. Calculate the general corrected death rates from the following information :—

Age–group	Town A		Town B	
	Population	Deaths	Population	Deaths
Under 5 years	3,000	80	6,000	150
5—20	15,000	140	1,4000	160
20—40	30,000	220	20,000	179
Over 40	7,000	300	4,000	180
Total	55,000	740	44,000	660

37. The deaths in two towns A and B are given according to the age groups and you are asked to state with reasons which of the two places you consider more healthy.

Age group	Town A		Town B	
	Population	Deaths	Population	Deaths.
Under 10	8000	180	20000	520
10–25	26000	130	100000	630
25–40	60000	450	250000	1980
Above 40	6000	400	25000	1770
Total	100,000	1160	395,000	4900

38. The following table shows the size of certain units sampled out in various coalfields of India. In which area is the average size of the coal-mining unit largest ?

No. of persons employed	Jharia No. of units	Raniganj No. of units	C. P. No. of units
Below 50	3	3	1
50–100	12	4	1
100–200	18	10	8
200–300	16	12	3
300–400	14	8	5
400–500	6	9	1
500–1000	19	13	6
1000 and above	13	9	7
	101	68	32

(1947–Mar.)

CHAPTER VII

DIAGRAMS

When a mass of statistical matter is collected it does not, in that state, serve much useful purpose unless it is properly classified as it is by the process of tabulation that the massy matter is put in logical order according to the units of measurement. But even with this device the power of visualization is partly realised. It is difficult even for an expert to bear in mind all the figures and in attempting to do so he is apt to lose sight of their relations. If tables are properly planned out and columns are provided for ratios and percentages this difficulty can be minimised, but if the matter is diagrammatically represented proportional relations between the various items can be easily and clearly brought out, the relations of amounts of space can be emphasized and above all they will aid us in holding in mind a large number of facts without much strain or effort. This is done by representing the matter by diagrams of one form or the other comprising of curves, geometrical figures or pictures to suit the particular case.

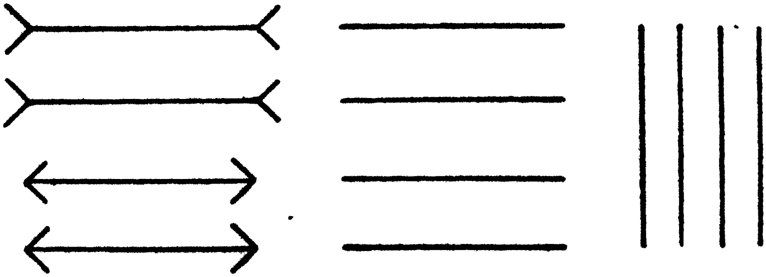
Diagrams bring outstanding features readily to the eye and thus present a picture of broad statistical facts based on tables but as such they cannot displace them. In fact it is very necessary that the tables, on which the diagrams are based should accompany them, as in the absence of such tables an *accurate* and detailed study of the problem in hand will be very difficult. They will, however, help us considerably in illustrating the tabulated data in a summarized form and thus render assistance in statistical studies by representing the matter with a suitable background.

Utility:—Though diagrams are not intended to add any new meaning to the existing statistical facts presented by them they help us ***in exhibiting the results more clearly***, particularly to persons who have not got the requisite mathematical training. Scanning the figures from the tables becomes very tiresome to the eye and it often confuses an average person while making comparisons which is very often the desired object. This can be easily achieved by presenting the facts by means of diagrams or charts which is ***the best way of appealing to*** the mind through the eye, through which medium man easily compares lengths, areas and volumes. When a manager of a factory has to review the productions of the various departments under his control it would help him considerably if the results are presented diagrammatically as in that case ***even at a glance*** he can have an idea of the progress. Similar will be the case of a busy physician who has to attend to several patients. Instead of recording their temperature and pulse in tables if adequate charts are maintained the charges will be ***readily intelligible at a glance*** and it will not only help him in diagnosing the case more efficiently but ***in saving a considerable amount of his valuable time and energy.***

Again a diagrammatic representation of facts will ***attract the attention*** more quickly and leave a more lasting impression on the mind of the observer. Instead of merely recording tables of Death Rates or of Illiteracy if diagrams are drawn, as is done in pictograms, they will be more effective. In the Glasgow Exhibition of 1938 incomes of labourers as well as of capitalists of coal companies in India were very effectively shown by diagrams and the observers were much impressed by the inequality and the miserable condition of labour in India. In short '***Pages of cold figures can be made to speak in clear tones if translated into the diagrammatic language.***' Above all with the help of graphic curves it is possible under normal conditions to forecast tendencies and likely changes in the future based on the experiences in the past. Looking to the growth of the population in the past one could, with reasonable accuracy extrapolate the population figure for the future. Even the missing data can be more easily interpolated with a fair amount of accuracy with the help of curves than by adjusting figures by formulas. It is a common experience of statistical workers that seasonal variations as well as correlation can be easily marked by a graphic diagram. In addition diagrams are of great use to those who do not want to go into details for want of time and yet want to have some information. Diagrams particularly of pictorial type evoke great interest which is almost impossible in the case of tabular statements. It is no wonder, therefore, that advertising pages of newspapers and magazines are often filled with picture, charts and diagrams designed to arrest the attention of readers and to convey quickly the essential ideas.

It must be borne in mind that a diagram represents only a broad picture and the accuracy of figures obtained from tables cannot be expected therefrom. It is very difficult to note the difference between 289 and 296 crores of rupees export when both are represented by small diagrams. Very often in the same diagram effort is made to show several things under review with the result that diagrams become confusing to the eye. This happens particularly when several curves are drawn in the same diagram. It is not unnsual to find the same data being represented differently by different parties having opposite interests and thus presenting quite different pictures. With the change of scales in the case of a graph the diagram will readily admit of diminution or of exaggeration of facts. If the space between two points on the horizontal scale is increased the curve showing the rise as well as fall will be materially flattened out, while with the decrease in space the rise will appear accentuated and the fall will look far more steep. Again, certain forms appear to be misleading even if they are geometrically correct. For instance, if two circles are drawn having diameters in the ratio of one to two, the area of the former circle will appear to be far less than half of the latter. If lines of equal lengths are drawn but with angles at both ends in different directions they will give an impression of having different lengths. Vertical lines appear to be longer than horizontal ones.

Lines of equal lengths appearing to be otherwise



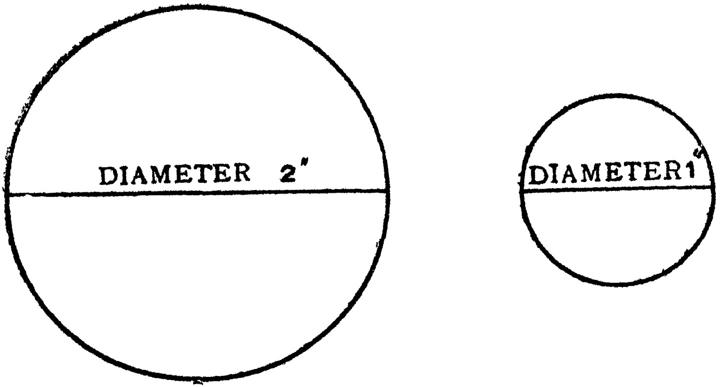


Fig. I.

However, if proper care is taken in drawing the diagrams they are very useful and that is why they are being more and more used in the statistical work.

Essential requisites:—In drawing a good diagram it is necessary to carefully bear in mind its artistic as well as scientific aspect. A diagram which has been designed with great care but has not been properly drawn and finished will fail to command the necessary attention with the result that the time and energy will be to a great extent wasted. On the contrary, the diagram drawn quite artistically but not correctly designed will receive undue attention and may even sometimes give wrong impression which is certainly far from desirable. It is needless to state that a diagram must be neat and *absolutely to the scale.* If productions of two mills are shown by lines or bars their lengths must be proportionate to the units of their production. It must have at the same time *a good visual* effect because if the diagrams are too complicated or if they are poorly designed they will fail to achieve the desired object. If in an area diagram too many items are shown or if many curves without proper distinguishing marks are drawn on a small sheet it would confuse the observer and the visual effect will be very poor. The design of the diagram should be simple yet impressive and it should not involve too much labour and yet should bring out comparisons wherever possible. Lastly the cost, time and labour involved should in no case be disproportionate to the utility. In short, maximum utility with minimum labour and cost should be the guiding principle.

Types of Diagrams:—There are various diagrammatic forms in which the statistical matter can be represented such as lines, curves, bars, circles, squares, spheres, prisms and pictures but while drawing a diagram the type giving the best result in the particular case will have to be carefully selected. The rainfall at a place for a number of years can be shown by means of lines, bars, curves, etc. but as the main idea will be to study the variations the use of a curve will be most suitable. If the density of population in the different provinces is to be depicted a statistical map would serve the purpose much better than any other form. In order to ascertain which type of a diagram will be most suitable for a particular purpose it is necessary to study the various kinds and their characteristics and for this it will be convenient to classify them into a few groups based mainly on their geometrical distinctions namely lines, areas and volumes. Curves as a rule will be in the Line Group, but for a matter of convenience and more because of a great importance attached to them in statistical studies they will be treated separately under the heading of Graphs. The various kinds of diagrams will be grouped as under :—

I. **One Dimensional Diagrams** i. e. line diagrams consisting of curves and lines as well as bars.

II. **Area Diagrams** i. e. *two* dimensional diagrams consisting of squares, rectangles, circles and angular diagrams.

III. **Volume Diagrams** i. e. *three* dimensional diagrams consisting of cubes, prisms, cylinder blocks whose volumes are in proportion to the figures.

IV. **Pictograms** e. g. statistical maps, charts and pictures.

Directions:—As already explained there are several forms of diagrams and naturally the teachnique of drawing each of them will differ but there are certain general directions which are applicable in almost all cases and so they have been given at the outset to avoid repetition.

1. *Heading*:—For each diagram a suitable yet a short heading should be given such as 'Bankers', 'clearing', 'Production of wheat in India', 'Income per Head.' But in case the matter is likely to be misunderstood by a short heading a sub-heading should also be given.

2. *Size*:—As a diagram is drawn to exhibit the statistical matter more clearly it is necessary that its size should be such as will allow its significant features to be clearly perceived. All the details must be clearly visible and the given data must stand in bold relief It should not appear clumsy and should not be too small or too big. It goes without saying that the size will also depend upon the space conveniently available.

3 *Scale*:—When the size of a diagram is fixed the scale to be adopted should be determined. This is written in the case of maps and charts, preferably in the right-hand top corner or in the left-hand bottom as 1 inch = 10 miles, or a line is drawn showing the scale as under

Scale
10 —— 20 —— 30.

In the case of curves the scale is shown on the vertical and horizontal lines. In determining the scale it should be borne in mind that the figure drawn should be in a position to show clearly the necessary details

4. *Drawing*:—As a diagram is required to be impressive it should be neatly as well as accurately drawn with the help of drawing instruments. The scale adopted must be strictly and rigidly followed. Colouring, dotting, crossing or cross-hatching should be done either in pencil or in colour but it should be free from blots. If lettering is done the letters should as far as possible read horizontaelly. If necessary a plain border should also be drawn.

5. *Guide or Key to a diagram*:—When in a diagram many items are to be shown it is done by distinguishing these by different types of colouring, crossing, cross-hatching or dotting. A key to such a diagram should be provided but in doing so it

should he seen that the visual impression conveyed by the diagram is not in any way distorted.

6. The original data on which the diagram has been based should be given, if possible, facing the diagram as this will help the observer to see the details with accuracy and there will remain no possibility of any misunderstanding or misrepresentation. If possible a reference to the source of the table should be mentioned.

In the end it must be remembered that the test of a good diagram depends upon the speed and ease with which the observer can interpret it and the cost and energy involved in its drawing.

1. **One Dimensional Diagrams:**—This group comprises of diagrams based on lines, straight as well as curve, but for the sake of convenience of treatment as stated already a separate chapter has been devoted to the study of Graphs.

Lines have no appreciable width and so a line-diagram is poor as far as its visual effect is concerned. It is to get over this difficulty that a bar diagram is preferred. It must be noted that the width has no mathematical importance and hence its size which should not be too big or too small, is governed mainly by the visual effect it is likely to produce. Bars may be simple or subdivided and further they can be placed vertically or horizontally.

Lines or Simple Bars:—When a line or a bar diagram is to be drawn the height taken is to be in proportion to the data and the scale should be fixed by taking into consideration the height of the largest bar and the size of the diagram. Lines or bars should be drawn as far as possible on a common horizontal line though it is not uncommon to find a vertical line being taken for this purpose. It must, however, be noted that the scale at the base should begin at zero, otherwise a false impression is given as will be seen from Fig. 4.

If at all it becomes very necessary to begin the scale at a higher figure than zero it should be shown by a false line as shown in Fig. 1 in the chapter on Graphs. The lines or bars should not be made to touch each other and there should be equal intervening space. The scale should be shown on the left hand side and not on the right as by having it on the left the observer knows beforehand what the scale is since he reads from left to right. In case the diagram is too wide it is better to give the scale on both the sides. In no case a diagram should look blurred or dirty and if space is less for bars lines should be drawn.

Table 1—Density or population per sq. mile.

Years	Bombay Presidency	Madras Presidency
1921	209	318
1931	235	350
1941	272	391

Density of Population

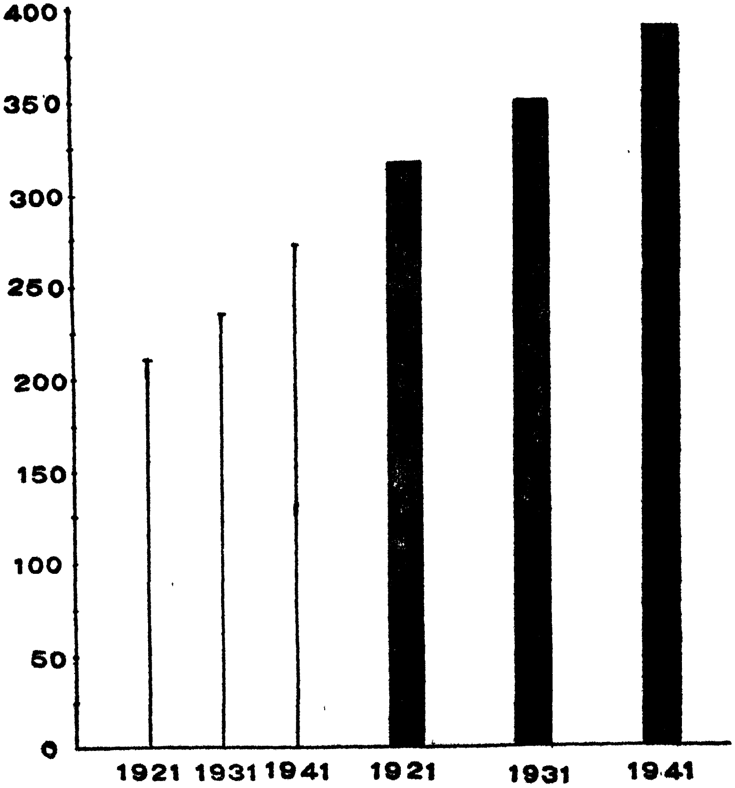


Fig. 2.

Density of Population in the Bombay Presidency

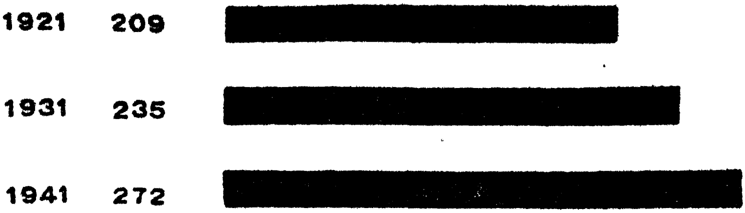


Fig. 3. Horizontal Bars.

In Fig. 2. the density of population in the Bombay Presidency has been shown by a line diagram while that of the Madras Presidency by a bar diagram. In Fig. 3. the density of population in the Bombay Presidency is shown by horizontal bars. It will be easily seen that a bar diagram produces a better visual effect than a line diagram.

Density of Population in the Madras Presidency

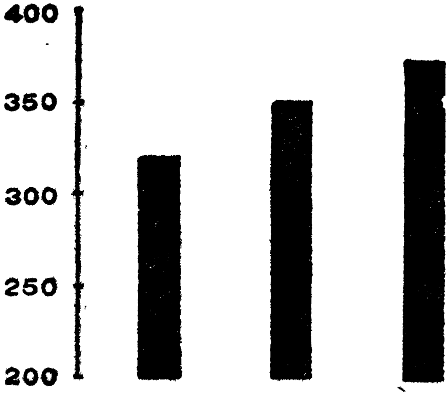


Fig 4. Bar-chart (*without the zero line*)

Density of Population

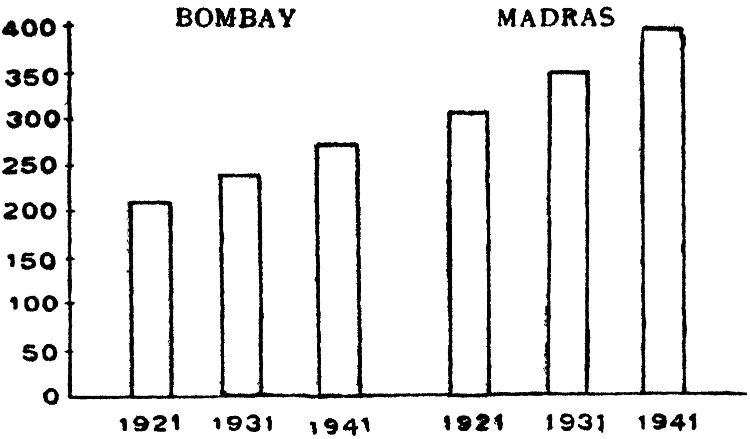


Fig. 5.

Density of Population

In case it is desired to effect a comparison it is better to place the two lines or bars side by side rather than at different places. In Fig. 5 the density of population in the two Presidencies has been shown by drawing the bars in two sets but in Fig. 6

Density of Population in the Madras Presidency

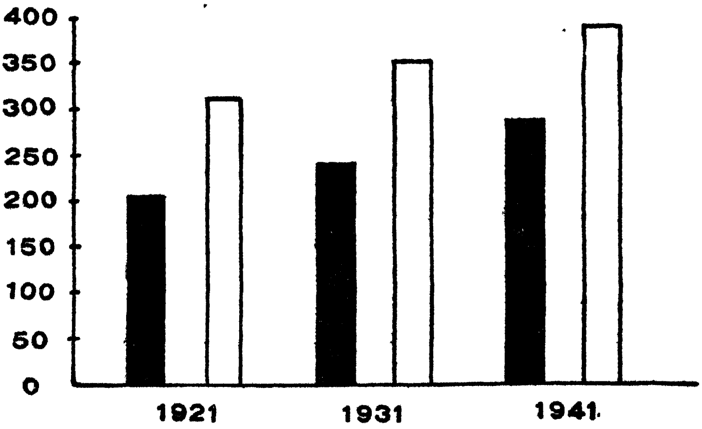


Fig. 6.

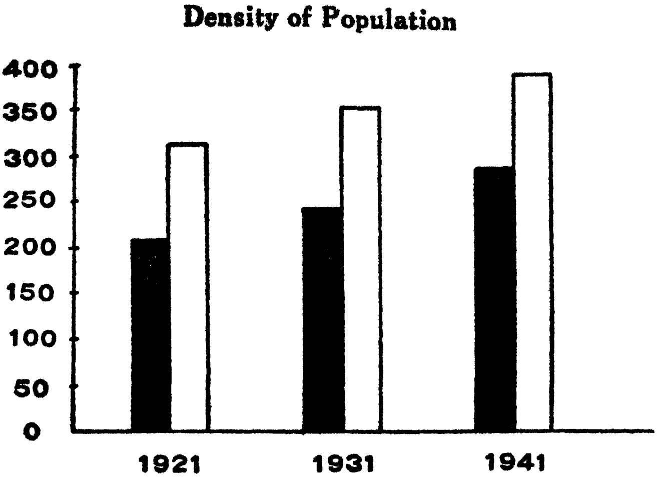
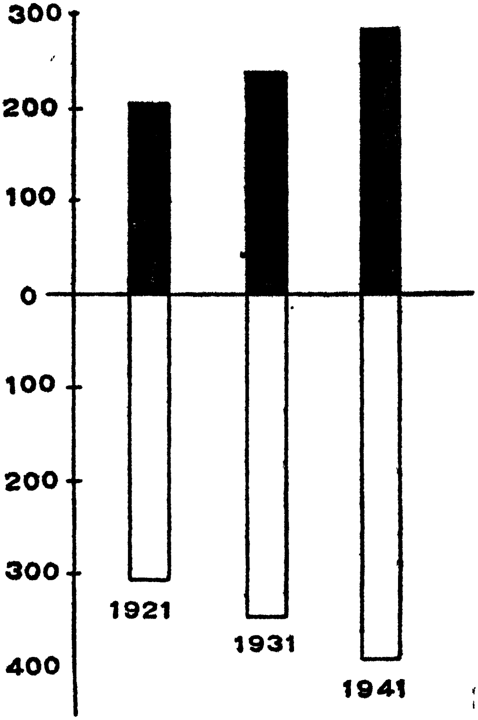


Fig. 7.

it has been done by placing those of the same year side by side and in Fig. 7 they are drawn touching each other. In such a case the two bars should be either differently coloured or there should be different crossings or cross-hatchings or dottings to distinguish the two bars. Charts prepared in this way are known as *Multiple or Compound Bar Charts.*

Duo-Directional Bars:—In all the above cases the bars had only one direction. They can also be drawn on both sides of the base and they are then known as Duo-Directional bars as will be seen form Fig. 8 and Fig. 9. In the former the bars have vertical directions while in the other they are drawn horizontally.

Bombay Presidency



Madras Presidency

Fig. 8. Vertical Duo-Directional Bars

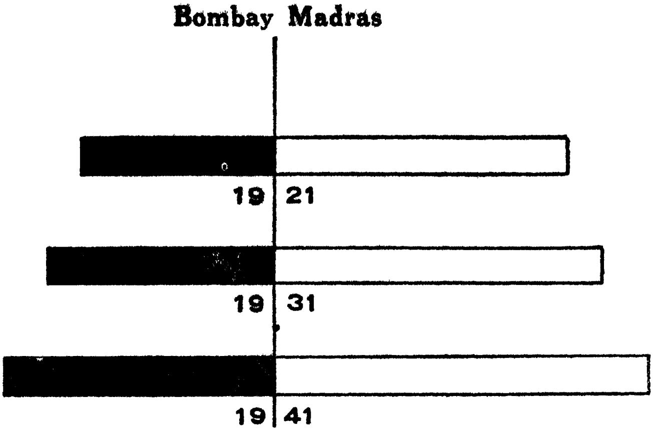
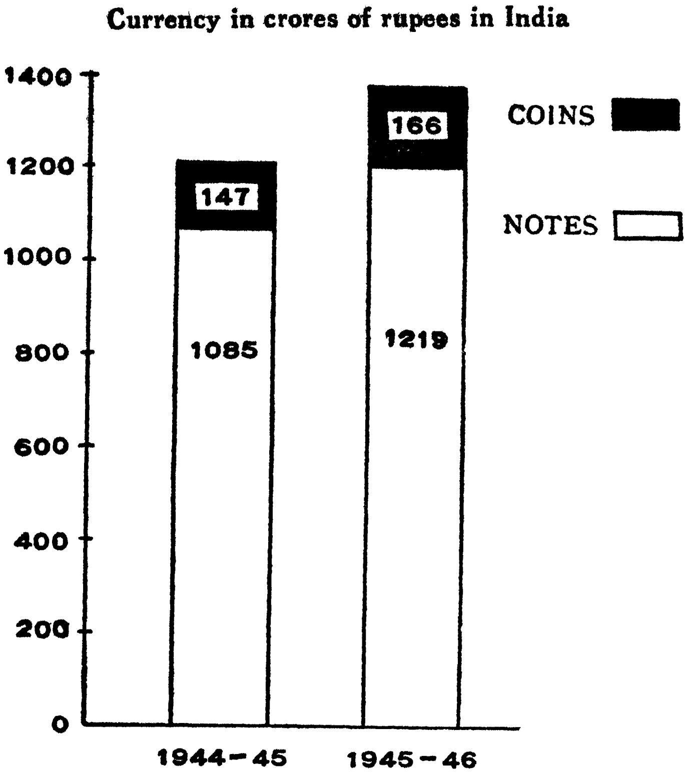


Fig. 9. Horizontal Duo-Directional Bars

Sub-divided Bar :—In the case of a simple bar the length represents the frequency or the magnitude but in a sub-divided bar diagram ,the bar is sub-divided into various parts in proportion to their values given in the data and the whole bar represents the total. The sub-divisions are distinguished by different colours, dots or crossings. The main defect in this case is that all the parts have not got a common base to enable us to compare accurately the various cases under consideration. Such charts are also known as Component Bar-charts.

Table 2—Currency in crores of rupees in India

	Notes.	Rupee Coin	Total
1944–45	1085	147	1232
Per cent	88.07	11.93	100
1945–46	1219	166	1385
Per cent	88.0	12.0	100



Currency in crores of rupees in India

Fig. 10. Divided Bar Diagram

Hundred percent bars:—This is one of the forms of percentage distribution diagrams and in this the component parts are drawn on percentage basis i. e. each part shows what percent it is to the whole. The data from table 2 has been represented in this form of diagram in Fig. 11. in horizontal bars but it can also be shown by vertical bars.

Currency in crores of rupees

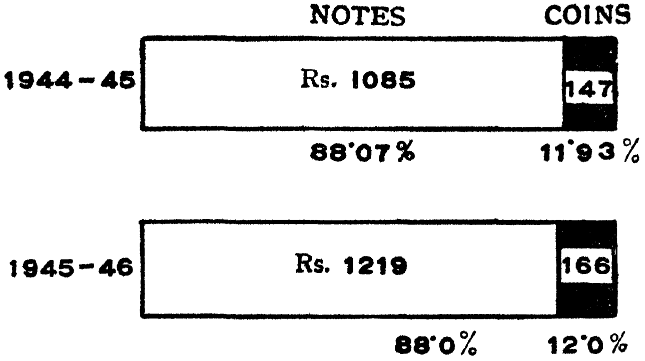
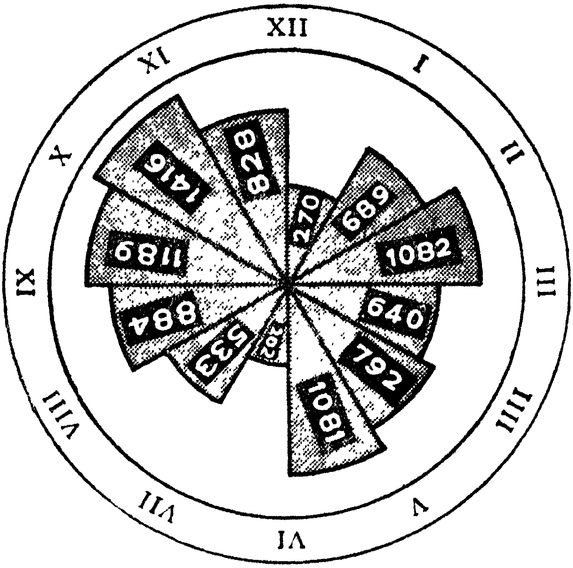


Fig. 11. Hundred Per Cent Bars

In area diagram pie or circular charts are made use of to show percentage distribution but for comparison such charts are not as useful as bar charts.

Polar Diagrams:—It is a form of bar diagram and is made use of to represent the number of events occurring in certain hours. Lines are drawn from a common point in various directions of different lengths proportional to the magnitudes or frequencies of the events. It appears like the rays of the sun or a clock but as the lines are drawn in various directions the comparison of their lengths becomes difficult as will be seen from Fig. 12 showing the number of cars passing by a particular road from 6 A. M. to 6 P. M. Due to this defect this form of diagram is not usually made use of.

Number of cars passing by Taj Road



Fig, 12. Polar Diagram

II. **Area Diagrams:**—In one–dimensional diagrams only the length is taken into account while the breadth is only for increasing the visual effect. In two–dimensional diagrams also called *Area* or *Surface* diagrams, the area represents the data and so the length and breadth have equal importance from the diagrammatic point of view. The most common forms of this type are rectangles and circles and they are again sub–divided into sections to represent different types.

Rectangles:—As the area of a rectangle is to represent the total units the base (i. e. breadth) should be fixed first and the length will then be ascertained by dividing the total units by the units of the breadth. If the component parts are to be shown their length will be similarly calculated.

Table 3. Clearing House Returns (in crores of rupees)

Years	Bombay	Calcutta	Rest.	Total
1944–45	186	196	58	444
1945–46	207	236	67	510

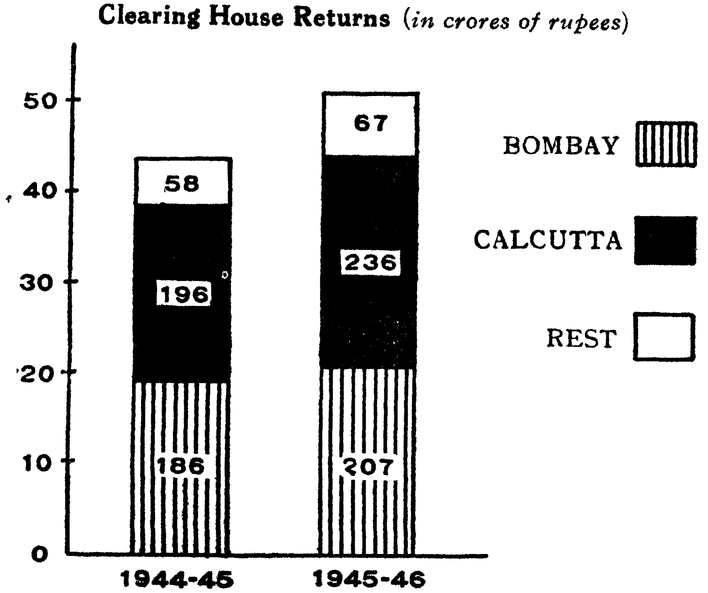


Fig. 13. Subdivided Rectangles

From Fig. 13 it will be seen that the first rectangle shows the Clearing House figures for 1944–45 with divisions showing amounts for Bombay, Calcutta and other centres and the other rectangle gives similar information for 1945–46.

Sometimes it is desired to show the actual figures as well as the percentages of the various items to the total. This can be easily done by taking the length as 100 and then adjusting the breadth to make up the required area. In this case comparison can be made on percentage basis as will be clear from the following example :—

Table 4—**Number of students in two colleges.**

Class	College A			College B		
	No.	Percentage	Cum. Percentage	No.	Percentage	Cum. Percentage
I Year	480	40	40	400	50	50
II Year	300	25	65	200	25	75
III Year	240	20	85	120	15	90
IV Year	180	15	100	80	10	100
Total	1200	100	...	800	100	...

In figure 14 the areas of the two rectangles show the total number of students in each institution and the parts show the strength of the classes. The veritical scale on the left shows the percentage strength of each class in the two colleges.

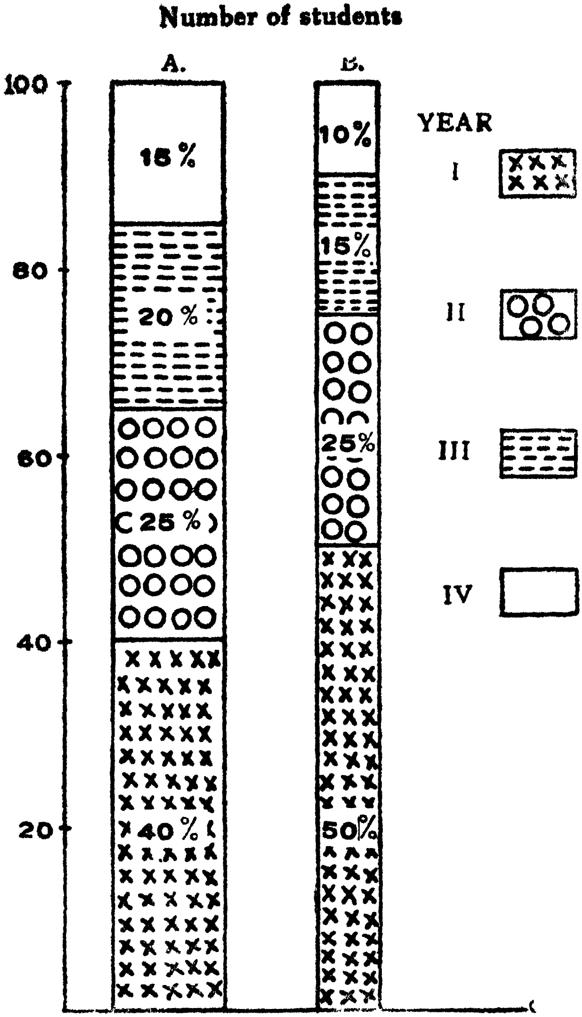


Fig. 14—Rectangles showing percentages

Such diagrams are not difficult to draw and are very useful particularly when comparison is to be made between two groups as well as between their components. In business percentages of cost, profit and loss can be very clearly shown by such diagrams. They are often known as ***Percentage Distribution Diagrams.***

Squares :—When the values of items to be represented diagrammatically have a wide range, it is difficult to represent them by means of rectangles because if the scale taken is a small one the heights of some of the recentagles will be disproportionately large and if otherwise, the item representing a small value will be too small. To get over this difficulty squares are drawn instead of rectangles as the sides of a square will be shorter as compared to that of a rectangle.

To draw square diagrams square root is taken of the values of the various items to be shown diagrammatically and then squares are drawn by taking a convenient scale as illustrated by the following example :—

Table 5.—**Exports from some of the Kathiawar ports**

Place	Exports in 1943–45 (in thousands of rupees)	Square root	Side of the square by taking 100 units to one inch
Navanagar	182,08	134.9	1.35
Morvi	40,67	63.7	0.64
Bhavanagar	6,67	25.8	0.26

Exports from Kathiawar ports

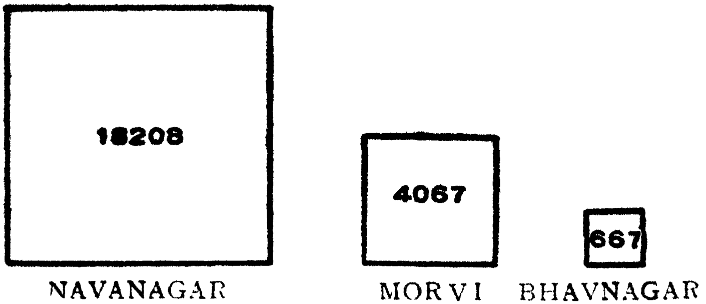


Fig. 15. Square Diagrams.

It will be seen that the square representing exports of Navanagar has a side of 1·35″ and an area of 1·82 sq. inches, the resulting scale, therefore, being one square inch equal to a crore of rupees.

Circles:—Circular diagrams or Pie–diagrams, as they are often called, are very useful in emphasizing areas besides being comparatively, easier to draw. With circles and sectors simple magnitudes as well as the total and component parts can be easily exhibited. Circles can be drawn by making either their diameters or areas proportionate to the magnitudes in question; but it is the latter method that is usually made use of and is particularly useful when magnitudes are to be shown by sectors. It is, however, very difficult to judge with precision the relative areas of circles. If two circles are drawn having areas in proportion of 1 : 2 the impression made will be far from reality (see Fig. 1) and this is the main defect in cases of circular diagrams.

Circular diagrams are normally drawn on the basis of the area and not on that of the circumference. In the case of a square its area is equal to the square of one of its sides. In the case of a circle it is equal to the square of its radius multiplied by 3.1416. It will be seen from this that the area of a circle will vary as the square of its radius.

To draw circular diagrams square roots of the data should be obtained and then these roots should be divided by some convenient number to obtain the radius as is done in the following example :—

Table 6—Cheques cleared in Clearing House in India.

Year 1	Crores of rupees 2	Square root 3	Radii in inches 4
1935	1937	44.0	.44
1940	2320	48.2	.48
1945	5902	76.8	.77

In the above table, column (3) gives the square roots of the amounts and the radii in inches have been arrived at by dividing them by 100 as otherwise the circles drawn will be too big.

Cheques cleared in Clearing Houses

(*in crores of rupees*)

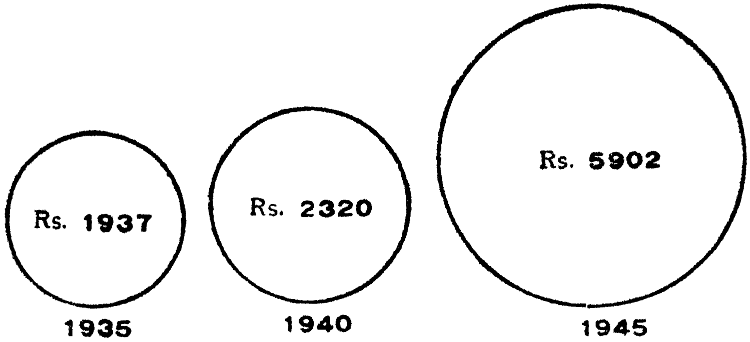


Fig 16 Circular Diagram

The area of a circle is equal to 3.1416 multiplied by square of the radius and so the areas of the three circles are 2.43, 2.89 and 7.44 sq. inches respectively representing 1937, 2320 and 5902 crores of rupees. The scale per square inch, therefore, works approximately to 800 crores of rupees to an inch.

Angular Diagrams:—As a bar or a rectangle is sub-divided to show its components, a circle can be divided into sectors and this can be easily done as the areas of the sectors are proportional to the angles at the centre. As there are 360° at the centre each degree of the angle at the centre represents $\frac{1}{360}$ area of the circle

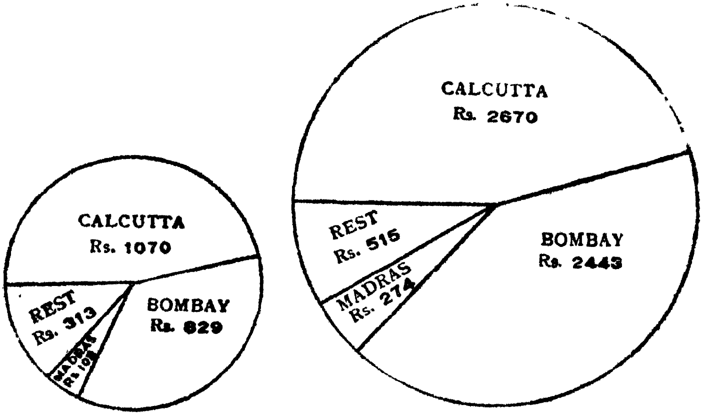
Taking this proportion, sectors can be drawn to represent the different component parts. As this working is comparatively simpler and as circles with sectors attract attention more easily, they are invariably preferred to squares.

Table 7—**Cheques cleared in India in Clearing Houses**

Centres (1)	1940		1945	
	Amounts in crores of Rs. (2)	Angles of the sectors (degree) (3)	Amount in crores of Rs. (4)	Angles of the sectors (degrees)
Calcutta	1070	166	2670	163
Bombay	829	129	2443	149
Madras	108	17	274	17
Other Centres	313	48	515	31
Total	2320	360	5902	360

Cheques cleared in Clearing Houses

(*in crores of rupees*)



1940 1945

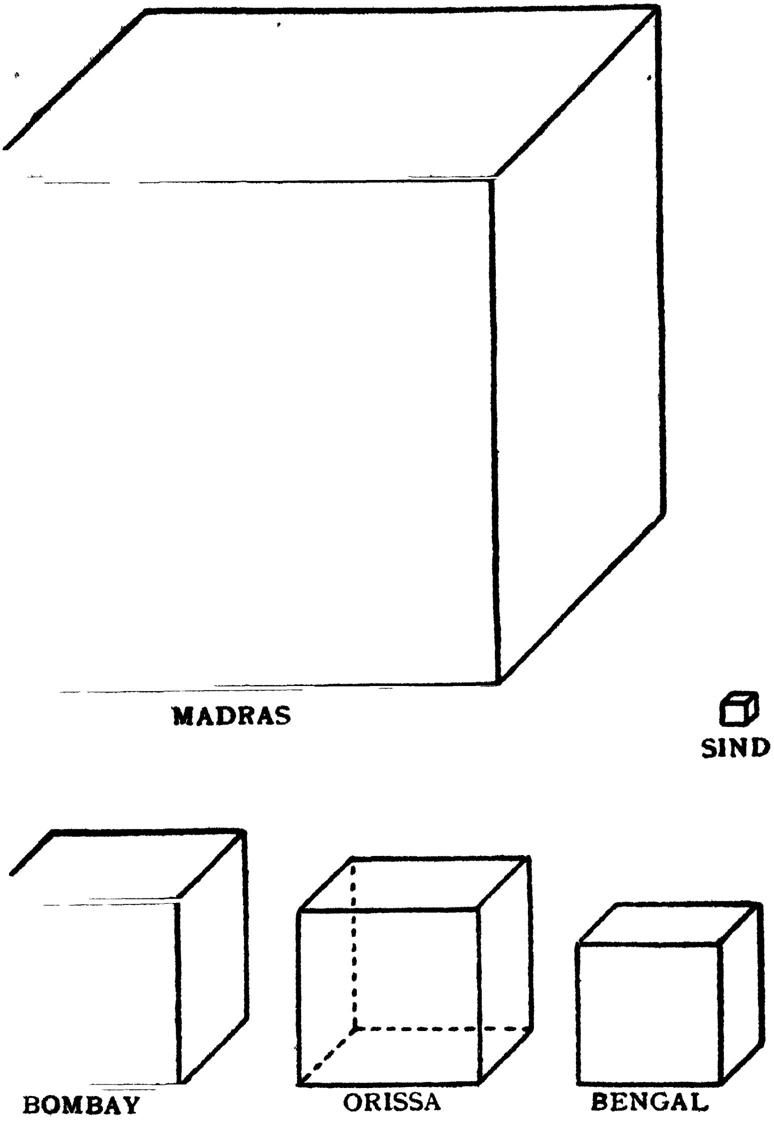
Fig. 17-Angular Diagrams

In figure 17 two circles are drawn to represent crores of rupees worth of cheques cleared in different clearing houses in the years 1940 and 1945, and their sub-divisions into sectors have been made on the principle that at the centre there are 360° and so the degrees for the various sub-divisions have been worked out in proportion to the total area. When the angles at the centre are marked with a protractor it is very easy to divide the circle into sectors having the desired areas. These sub-divided circles present a considerable difficulty when the area is to be ascertained particularly as compared to the calculation of the areas of subdivided squares or rectangles. It is again inconvenient to read the names given to the various sectors, question of comparing areas of the same type being still more difficult.

III. Volume Diagrams:—Volume or Three-Dimensional diagrams comprise of cubes, spheres, prisms, cylinders, and blocks. Of these, cubes are mainly used and their sides are drawn in proportion to the cube roots of the magnitude of the data. They are particularly useful when the data has a very wide range. In such a case it would be very difficult to represent the quantities even by squares not to think of the bars or rectangles as the figures drawn will show a very great divergence.

Table 8—**Area under cocoanut in 1941–42.**

Province 1	Acres 2	Reduced figures 3	Cube roots	Side of a Cube
Madras	596147	20557	27.40	2.74
Bombay	28547	984	9.95	1.00
Orissa	28425	980	9.93	0.99
Bengal	12700	438	7.59	0.76
Sind	29	1	1.00	0.10



Area under cocoanuts in 1941-1942

Fig 18 Volume Diagrams (cubes).

In table no. 8 in column no. (2) acres under cocoanuts are given according to provinces but the range is very, very wide. The acreage in Madras is 596147 while that in Sind is only 29 and so rectangles or squares cannot be conveniently drawn for this data. As the figures of acreage are too big, they have been reduced by dividing each of them by 29 and writing them in column (3). Even the figures of cube roots are too big to be taken as sides of cubes and so each of these figures has been reduced by dividing it by 10. With the sides given in the last column the cubes are constructed to represent the various provinces. For this purpose, a square with the given side is drawn and then another square of the same dimension is placed behind it at a convenient distance and the corresponding corners are then joined as shown in the diagram for Orissa in figure 18.

IV. **Pictograms**—Maps and Pictures.

Tabular data can also be represented by pictograms, cartograms, maps and pictures as these methods help considerably in attracting the attention to the statistical matter which when presented in the ordinary diagrammatic form is very often ignored.

Pictograms are diagrams of pictorial or semi-pictorial nature and are drawn in different sizes according to scale to demonstrate the tabulated matter. For instance, if 'Per Capita Deposits' of two countries are in the ratio of 2:1 two symbols of pouch holding money can be drawn the proportion to their seizes being 2:1 to maintain the proportion to the data. Similarly, if imports of two countries or of two years of the same country are in the propertion of 1:3 two ships full of cargo will be drawn with sizes in the given proportion i e. 1 : 3 and the names of the countries or years will be shown to designate them. The other method is of maintaining the size but making a change in the number of figures drawn. In the case of 'Per Capita Deposits' in this method two pouches will be shown for one country and one for the other country thereby showing that the 'Per Capita Deposits' in the two countries are in the proportion of 2:1.

The method of Pictorial Statistics is credited by some to Dr. Otto Neurath, Director of Social and Economic Museum, Vienna, and is, therefore, also termed as **Vienna Method or Isotype**

Method' a word formed by taking the initial letters of International System of Typo-graphic Pictorial Education, (Isotype).

Table—9, **Density of population per sq. mile.**

Country	Density	Symbols to be shown by taking 41 as units
United Kingdom	685	16.7
Belgium	654	16.0
Japan	443	10.8
India	195	4.8
U. S. A.	41	1.0

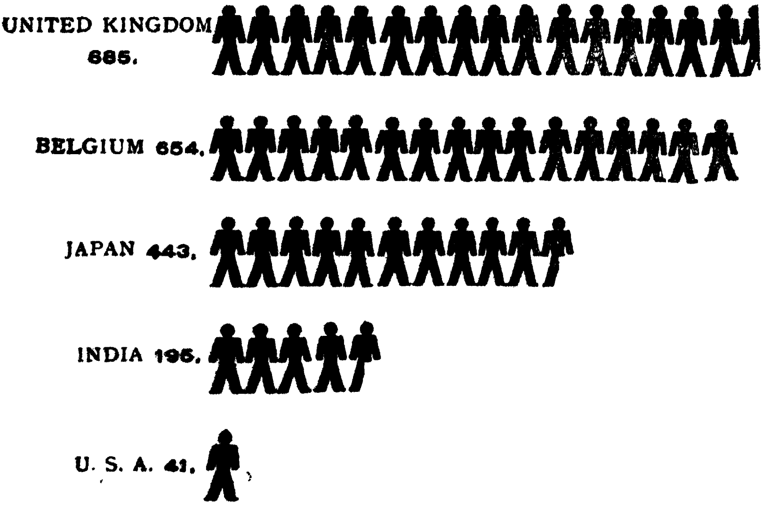


Fig. 19.

Pictograms are very useful in attacting attention as well as for showing comparisons, They are easily understood and create a more vivid as well as lasting impression on the mind than a table. People very often pass over tables but pictorial diagrams are rarely ignored even by disinterested readers. It must, however, be accepted that they are sometimes misleading and they cannot be made use of when a complicated data is to be represented. From the accuracy point of view their utility is much less and they add almost nothing to the existing tabular data.

Mapgraphs, Cartograms and Statistical Maps are widely used to represent special distribution such as areas under different crops, population-its density and classification, sales distributed over the different sales-territories. On the map of the territory in question the magnitude of the data is shown (a) by points, dots or crosses, (b) by writing the actual amounts or (3) by deepening the colour in proportion to the magnitude of the data or by using different colours or by cross-hatching and it is indicated by the key shown in one of the corners of the map.

(a) **Points or Dots** :—On a map in the areas concerned points or dots (sometimes crosses also) are marked, their numbers depending upon the magnitude of the data. If in the three areas A, B and C with 600, 400 and 200 units are to be shown on a map it can be done by drawing three circles with 6, 4 and 2 dots in the respective areas of the map.

The second method is of putting dots in these areas which will be of sizes proportionate to the data. The dot in area A will be three times in size of that of C area and the size of the dot in B area will be double in size of the dot in C area. In short, the sizes of the three dots will be in the ratio 3 : 2 : 1. Yet, there is the third method in which all the dots are of the same size but shading is done in each case which will be in proportion to the magnitude of the data. The dot in area A will be completely shaded while the one in area C will be shaded only one third of its area.

(h) **Writing Numbers**:—The easiest method is of writing the actual amount given in the data in the respective areas. This method is simple but less attractive.

(c) **Colouring or Cross-hatching**:—Colour can be used to indicate the relative density. It can be made more and more deep to show higher and higher magnitude. This method is used in physical maps to show the depth of sea or heights of mountains, Instead of this, different colours are used to show different magnitudes and this is shown in the key, given in one of the corners of the diagram. Instead of this, cross-hatching is also done to indicate densities. The higher the magnitude of the data the closer will be the lines of cross-hatching.

Pin-maps—All the above methods are not convenient where the data goes on changing fast and yet the diagrams are required to be kept upto-date. In such cases Pin-maps are most suitable. In this, large scale maps are mounted on card board or cork and specially prepared pins with short shanks or legs and with round glass heads of different colours are struck on the map to indicate the required information. The index to the coloured heads of the pins is given on the map to facilitate the work of interpreting the data.

Exercises.

1. Describe the importance of diagrams in statistical work.
2. Name the important types of diagrams and describe in detail the method to be followed in drawing any two of them.
3. What precautions will have to be taken in drawing a good diagram? What in your opinion, are the tests of a good diagram?
4. Differentiate between
   (a) line and bar diagrams,
   and (b) area and volume diagrams.
5. Write short notes on :—
   Sub-divided bar diagram, polar diagram, angular diagram, Isotype method and pin-maps.
6. What is the utility of area and volume diagrams in statistics? Outline the method of constructing a volume diagram.
7. Of all types of diagrams pictograms are most useful. Support or criticise this statement with suitable illustrations.
8. What mistakes are generally found in a diagram and how would you guard against them?

9. **Represent the following diagrammatically :—**
Export of bales of wool from Bombay to various places.

	1937–38	1938–39
Liverpool	16762	24487
Philadelphia	4158	4747
Hill	1703	1450
Hamburg	1048	657
London	641	540

(1940–Mar.)

10. Show by diagrams the distribution of the Reserve Bank shares as given below :

	No. of Shares in thousands.		No. of shareholders in hundreds.	
	1937	1938	1937	1938
Bombay	201	206	215	208
Calcutta	125	123	145	131
Delhi	94	93	157	149
Madras	60	59	91	87
Rangoon	19	18	18	16

11. Represent the following diagrammatically :—
Infant mortality in different towns.

Bombay	274	London	66
Calcutta	274	Berlin	82
Madras	251	Paris	93
Nagpur	323	Oslo	23

12. Draw an angular diagram to show the following :—
Industrial Workers in India.

	Men	Women	Children	Total
1914	746,773	144,157	60,043	950,973
1936	140,0210	239,875	12,062	165,2147

13. Illustrate the following by suitable diagrams ;—
(a) Production of sugar in India.

Year	In 1000 tons
1929–30	311
1932–33	645
1935–36	1087

(*b*) Number of motor cars imported from U. K.

1927–28	3600	1933–34	5348
1930–31	2885	1936–37	6337

(*c*) Number of looms and spindles.

	Looms	Spindles
Bombay Island	66,753	28,89,509
Ahmedabad	48,833	18,76,370
Sholapur	6,678	2,86,004

14. Design a single diagram to exhibit the entire information given by the following statistics, so as to bring out how the causes of death are distinguishable in their intenisities and how they are also definitely associated with social status :—

Cause of death	Illegitimate children	Poorest classes	Unskilled workers	Artisans	Well-to-do classes	All classes
Diarrhoea and Enteritis ...	24	17	11	7	1	60
Prematurity and atrophy ...	56	33	32	34	24	179
Bronchitis andPneumonia ..	22	25	19	11	4	81
All causes ...	102	75	62	52	29	320

(1939–Mar.)

15. Show by means of circular diagrams the following information regarding the number of female scholars in recognised institutions in India :—

	1930–31	1935–36
In Arts Colleges	2,744	5,329
In High Schools	85,879	1,23,949
In Middle Schools	1,70,897	2,16,010
In Primary Schools	19,81,549	25,05,077

16. Show by means of a three dimensional diagram the production of salt in India in 1935.

	Rs.		Rs.
Aden	19,81,299	Burma	531,009
Bengal	1,079	Gwalior	4,725
Bombay & Sind	3,102,656	Madras	2,289,790
		Northern India	3,778,579

17. Draw square diagrams to show the Indian population in 1936 in the following countries :—

Ceylon	659,311	Kenya	38,325
Mauritius	268,211	Fiji Islands	85,002
		Transaal.	25,561

18. Draw Bar diagrams to illustrate the following :—

Number of motor vehicles in states in 1936.

Hydrabad	3609	Travancore	1583
Mysore	2687	Gwalior	1216
		Jaipur	776

19. Show by means of a suitable diagram the following :—

(*a*) Principal heads of Revenue of the Government of India in 1937–38.

	Rs. in lakhs		Rs. in lakhs
Customs	4393	Taxes on income	1212
Central Exercise Duties	774	Salt	835
Corporation tax	150	Opium	46
		Other heads	100

(*b*) Expenditure in the year 1932–36

	Rs. in lakhs		Rs. in lakhs
Calcutta	319	Madras	32
Bombay	264	Karachi	64
		Chittagong	7

20. **Represent the following diagrammatically :—**

Table showing the number of workers in thousands occupied in some of the important industries of India.

Industries		1921	1931
Textile	...	4031	4102
Wood	...	1581	1632
Metals and Ceramics.	...	1811	1738
Food	...	1653	1478
Dress and Toilet.	...	3404	3381
Total	...	12480	12331

(1940–Oct.)

21. With the help of diagrams facts can be misrepresented. Discuss.

22. How far the statistical representations of facts by diagrams is more advantageous than by giving statistical tables ?

23. The following table gives the details of monthly expenditure of two families A and B. Show this by means of hundred per cent bars.

Items of expenditure	Family A Income Rs. 500 p. m.	Family B Income Rs. 800 p. m.
1. Food	140	240
2. Clothing	80	160
3. House Rent	100	120
4. Education	30	80
5. Fuel and Lighting	40	40
6. Miscellaneous	40	80

CHAPTER VIII

GRAPHS

Line diagrams comprise of two sub-groups ; (1) those drawn with the help of straight lines and (2) those consisting of curve or broken lines. In the preceding chapter diagrams falling in the first sub-group have been discussed, reserving the discussion of the other group for a separate chapter, because of its wide applicability as well as of great importance attached to it in statistical studies.

'A curve is a line no part of which is straight' while a broken line is a combination of straight lines. Statistical 'Curves' are seldom real curves but more often broken lines. Curve charts or graphs may represent data of time series i.e. pertaining to course of events extending over successive periods of time or of frequency distribution showing a state of affairs existing at a given period of time. Charts of the former type are called *Historigrams* while those of the other, *Histograms*. Again graphs can be constructed on *Natural Scale* or on *Logarithmic Scale.*

Utility :—Graphs form a convenient medium for the symbolic expression of numerical data and as such particularly to an untrained person a curve serves as a pictorial argument appealing to reason through the eye far more than a statistical table, though it is likely to supply more detailed as well as accurate information. Generally, it is difficult for an ordinary person to remember figures from statistical tables to make comparisons, but this difficulty can be easily got over by presenting the data by curves which will make comparisons self-evident. It is not easy even for a doctor to grasp the trend of temperature of a patient, ailing for several days if it is recorded in a tabular form but when it is represented graphically even a layman will be able to mark the changes without much effort as well as loss of time, a factor of a very great importance for a busy man. With the help of a curve-chart forecasting becomes far simpler as the trend of the curve in the past is sure to give under similar conditions an indication for the future. If sales figures of some years are graphically represented one can with reasonable accuracy estimate the figures for the future unless, in the mean time, circumstances have changed materially. For recording historical changes in different periods this is a suitable method and more so as it supplies basis for finding out by interpolation the missing value of a variable. Very often by looking to graphic representation new ideas particularly regarding correlation or connection between two different factors or events arise in our mind and to research workers this gives considerable help. It is comparatively easier to construct a graph and it also facilitates the finding of mode, median and the quartilies. In the words of Boddington *'The wandering of a line is more powerful in its effect on the mind than a tabulated statement. It shows what is happening and what is likely to take place, just as quickly as the eye is capable of working'.*

It must however, be noted that curve charts or graphs are not free from defects. The main difficulty is that graphs cannot be quoted as can be done in the case of numerical figures. They cannot also be used in cases where very great accuracy is demanded. Numbers can be entered in tables to any place of decimals and they can be read as such but from graphs this is not possible particularly when the graph is required to represent big numbers. The main defect in this method is the possibility of deliberate misrepresentation of facts. By taking varying scales for drawing a graph, figures obtained are apt to present different pictures. With the change of a scale the curve can be made to appear more flat or more steep and particularly a layman is bound to carry a wrong impression. In spite of all this, Graphic statistics do play a role of its own. They cannot altogether displace numerical statistics but the utility of the data will certainly be increased by this method of representation.

Construction:—Usually a graph is drawn on a specially ruled paper with faint vertical and horizontal lines—known as co-ordinate axis—intersecting at right angles. The point of intersection is called the origin and is shown by zero point. The horizontal line is called the x-axis or the axis of abscissa while the vertical line is designated as the y-axis or the axis of the ordinate. The position of any point is fixed by measuring its distance from the axis of X and that of Y. Distances measured upwards of the Y-axis from the origin are positive while those downwards are negative. Similarly, distances on the right side from the origin on the X-axis are possitive and those on the left side are negative.

This system of graphics in statistics is said to have been introduced by William Playfair of Scotland who published his Atlas of Charts in 1786 but at present it is far more in use in U. S. A. than in England or Scotland. A number of associations led by the American Society of Mechanical Engineers appointed a joint committee in 1915 to consider Standards of graphic presentation after the First Great War, but it was superseded in 1926 by the Sectional committee with Mr. Willard C. Brinton as the Chairman. The committee made suggestions, the more important of which, are incorporated in the important rules given below.

1. The general arrangement of a diagram should always proceed from the left to right and the ordinate and the abscissa should be constructed to read from bottom to top and from left to right respectively. The Lorenz curve is an exception to this rule.

2. As far as possible, quantities should be represented by lineal magnitudes as areas or volumes are likely to be misinterpreted.

3. It is always convenient to have the data side by side as it will be of help for more accurate and detailed study.

4. The scale chosen should be such as will conveniently include all the observations without making the curve appear too steep or too flat. This will no doubt depend largely upon the space available as well as on the extent of the magnitude of the data. Equal distances anywhere on the horizontal or vertical scale must represent equal values but the same scale need not in special cases be taken in both the cases.

5. The zero line of the scale should be sharply distinguished from other co-ordinate lines and that of the vertical scale should normally appear in the curve diagram. If this is impracticable this zero line should be indicated by a horizontal break by means of a false base line as shown in figure 1.

6. When the scale of a diagram refers to dates or years this ***independent variable*** should be plotted on the axis of x while the ***dependeut variaable*** should be on the verticle axis.

7. In curves representing a series of observations, points should be sharply distinguished from the ruling and wherever possible it is advisable to show clearly in the diagram all the points representing the different observations.

8. Figures showing the scale taken for the diagram should be written on the left hand of the vertical scale and at the bottom of the horizontal scale.

9. All lettering and all figures in the diagram should be so placed as can be read easily from the base at the bottom or from the right-hand edge of the diagram.

10. The title of a diagram should be as clear and complete as possible. Sub-titles or description should be added if necessary to ensure clearness.

11. When it is required to draw different curves in the same diagram it is necessary to differentiate them by (*a*) different coloured lines, (*b*) lines differently characterised e. g. thick, thin, dotted, crossed, continuous, broken or (*c*) simple curves with caption or designation showing the nature of the data represented. In the first two cases it is necessary to give the 'Key' or 'Legend' on the chart to distinguish the curves.

12. If curves are drawn to show comparison, the same scale must be taken in all the cases to effect proper comparison. The other way of making a comparison is to use transparent paper for drawing the curve and then placing this drawing on the first diagram.

13. Where curves are drawn on logarithmic co-ordinates the limiting lines of the diagram should each be at some power of ten on the logarithmic scale.

False Base Line. ;—If fluctuations are relatively small to the size of the variables they cannot be given prominence unless the vertical scale is sufficiently amplified. But in this case if the scale is to be shown right from zero line of the ordinate the height will be too long to be conveniently accommodated in the diagram. This difficulty can be easily got over by not showing the vertical line completely. As the fluctuation will be limited to a narrow range a large part of the ordinate at the bottom will be without plotted points and so this part can be easily left out without any inconvenience. As it is necessary to show that there has been a break in the continuous vertical scale beginning from zero it is done by a double saw-tooth line as shown in (Fig. 1.) False base line should not be normally used as such a diagram is likely to create a wrong impression. Some make use of this device to save space and to effect paper economy but this is not a good practice as this will result in making the fluctuations look more than they really exist.

Table 1.—**Index Number of wholesale prices.***

Average of weeks	Agricultural commodities.	Industrial raw material.
1946		
March	295.7	283.5
April	294.3	282.2
May	299.3	288.9
June	302.8	299.7
July	303.8	296.1
August	303.5	297.0
September	302.2	297.9
October	321.7	303.5
November	336.0	354.0
December	326.0	348.5
1947		
January	314.6	343.3
February	324.8	352.1

*Reserve Bank of India Bulletin—June, 1947.

Index number of wholesale prices

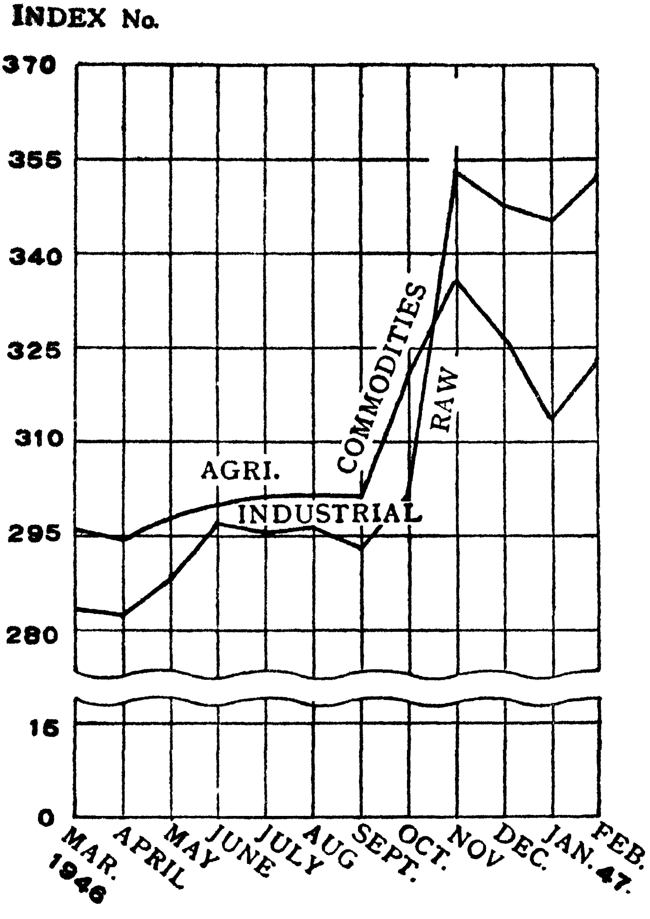


Fig. 1.

From the foregoing diagram showing the two curves with captions, 'Industrial Raw Material,' and 'Agricultural Commodities' it will be seen that the vertical scale is not continuous but broken. By this device of *false line* it has become possible to represent the relatively small fluctuations in a significant manner. This has been achieved by dropping some portion of the vertical scale as there are no plotted points upto 280.

To avoid possibility of a wrong impression a false base line has been shown in the vertical scale.

Frequency Distribution Graphs

Graphs are drawn to represent frequency distribution as well as time series showing continuous changes in statistical data over a period of time. Frequency distribution can be presented in a tabular form as shown in the table giving weights of hundred persons but if the same data is represented graphically, sizes of the number can be more readily and easily seen.

Table 2—**Weights of 100 persons**

Weights.	No. of workers f	Cum. frequency cf.
100 – 110	8	8
110 – 120	11	19
120 – 130	12	31
130 — 140	17	48
140 — 150	19	67
150 — 160	10	77
160 — 170	8	85
170 — 180	7	92
180 — 190	5	97
190 — 200	3	100
	100	

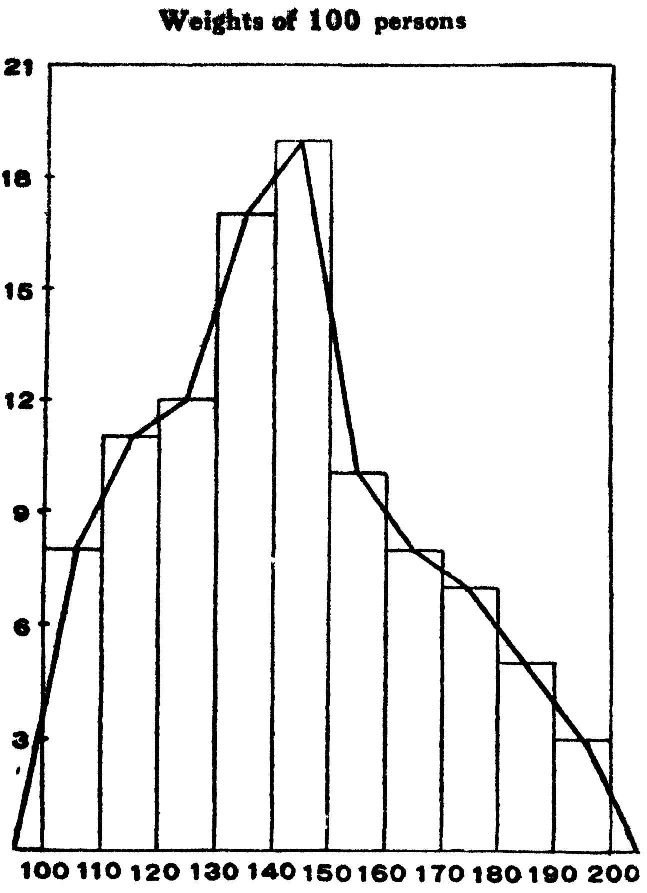


Fig. 2. Histogram and Frequeuey Polygon.

For this type of graphic representation the vertical scale is used for representating frequencies of the various classes the magnitudes of which are shown on the horizontal scale. After choosing convenient scales for the horizontal and vertical lines, points are plotted taking the points of the frequency and the magnitude i. e. mid-point in each case. The reason for taking the mid-point is that as the number of observations for the particular class is likely to fall somewhere within the two limits of that class it is in fitness of things to take the average

i. e. almost the mid-point in the absence of the actual information. In frequency diagrams it is not necessary to show zero position if it is not convenient to do so.

When the points are plotted the figure is drawn in one of the following ways :—

(1) By means of a rectangular diagram, called *Histogram* ;

(2) By means of a *Frequency Polygon*, or

(3) By means of a *Frequency Curve.*

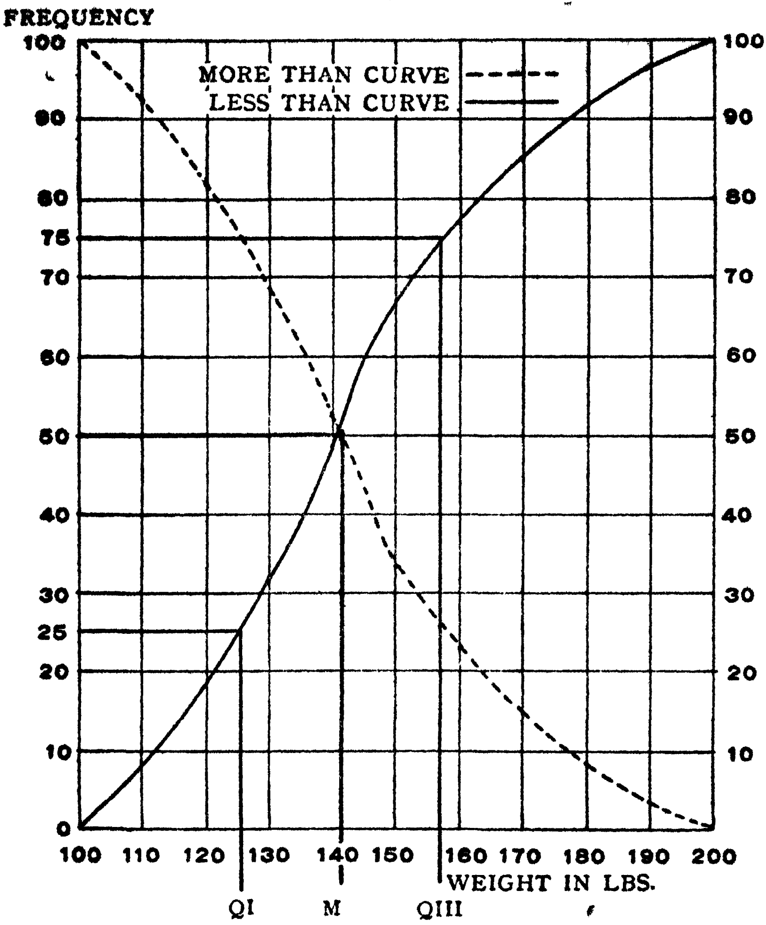
Histogram :—Histogram, Block diagram, also called Staircase diagram is formed by drawing rectangles on the horizontal line one over each class-interval as shown by the bold lines in Fig. 2. The area of each rectangle is equal to the frequency of that class on which it is drawn while the total area covered equals the total frequency units. The heights of the various rectangles will easily give a better and accurate impression to the eye of the distribution of the frequencies of the different classes that can be obtained only by scanning the table. With the help of the histogram the frequency curve can be drawn with greater ease than can be done with the help of a frequency polygon and it is particularly convenient for representing unequal group intervals. The main defect of this diagram is that if the class-intervals are changed the figure formed will be different. This defect can be remedied by drawing a frequency polygon or a curve instead of a histogram. Sometimes the available data is found to have classes of unequal intervals but it can be represented by constructing histogram in the usual manner without much difficulty.

Frequency Polygon :—This figure is formed by joining the plotted points by straight lines as shown in Fig. 2. and its area is approximately the same as that of the rectangles in the case of a histogram. However, the question of the area is not given prominence in this type but it is very useful for locating the mode. This type of a diagram can be easily understood even by a layman and more than one polygon can be easily drawn in the same diagram which is not quite convenient in the case of a histogram. But the main defect of the frequency polygon is that it shows somewhat sudden changes and a picture of uniform and regular variation in the magnitude is not obtained. This is possible in the case of a *Frequency Curve*.

Frequency Curve:—This curve is obtained by smoothing out of the minor irregularities of a frequency polygon bearing in mind that the area enclosed should represent as far as possible the total amount of frequencies in the entire distribution. The best and easy method is first to construct a histogram and a frequency polygon and then to draw with free hand a smoothed curve touching the horizontal line and then passing slightly over the apex of the polygon and again touching the base line thus forming more or less a bell-shaped figure. This is also called *Normal Frequency Curve or the Normal Curve of Error.* As the curve in the course of construction undergoes smoothing by rounding of angles it is also known as *Smoothed Histogram* or a '*Smoothed Frequency Curve*'

Cumulative Frequency Curve:—This curve is obtained by plotting the cumulative frequency instead of the simple frequency. It runs more regularly than the ordinary frequency curve without being affected by variations in group intervals and is particularly useful for finding out median, quartiles, deciles as well as percentiles. The mode can also be located on this ogive by finding the steepest part of the curve. But this is not as easy as locating a median. This curve is also smoothed out if necessary. As it assumes the form of an arch it is called 'Ogive' or 'Ogee'.

It will be noticed from Fig. 3 that the curve has been drawn by plotting the points by taking the cumulative frequency as given in the last column of table No. 2. The point to be particularly noted in laying out an ogive is that in the case of a cumulative frequency curve the plotting is to be done *at the uppermost limit* of the class interval and not on the mid-point as is done in case of a frequency polygon.



Weight of 100 persons.

Fig. 3. 'More than' and less than' curves.

'More than' 'Less than' diagrams:—These are the different forms of cumulative frequency graphs and they can be obtained by arranging the cumulative frequency tables to suit the particular case as will be seen from tables nos. 3 and 4 as obtained by arranging the matter in table no. 2 giving simple frequencies of weights of 100 persons.

Table 3.—**Cumulative frequency table 'More than'**

Weight in lb.	Number of persons
More than or above	
190	3
180	8
170	15
160	23
150	33
140	52
130	69
120	81
110	92
100	100

Table 4.—**Cumulative frequency table—'Less than'**

Weight in lb.	No. of persons
Less than or under	
110	8
120	19
130	31
140	48
150	67
160	77
170	85
180	92
190	97
120	100

It will be seen from the two tables that the data given in table no. 2 has been re-arranged to obtain the two tables from which 'More than' and 'Less than' curves are drawn as shown in Fig. 3. If from the point of intersection of these two curves a perpendicular is dropped to the base the median value can be ascertained by measuring the length of the axis of X from the point of origin to the foot of the perpendicular. Sometimes instead of plotting the numbers of cumulative frequency the whole data is turned into percentage to the total and it is then plotted as this affords better facility for making comparisons.

Time Series Graphs

'Time Series' also called 'Historial Series' gives the numerical record of changes of the variable in a particular period taken at successive intervals of time. The graph obtained by plotting the historical data is called 'Historigram' and it should not be confused with 'Histogram' which is obtained by drawing rectangles in the case of a frequency diagram as shown in fig. 2.

In the case of historigrams the horizontal scale is used to indicate intervals of time while the vertical scale is for measuring the strength or magnitude of the phenomena occuring in the different times of the period. Great care is necessary in choosing scales both vertical as well as horizontal to guard against exaggerating the features of the data by making the curve appear too steep or too flat. Sometimes it is necessary to draw the graph of a series of component parts as shown in Fig. 12 by band curves. In this case the scale taken should be such as will show the component parts distinctly and the different parts can further be distinguished by different colours or shadings.

Time series graphs are drawn to study the changes over a period of time of (1) a single variable or of (2) two or more variables. These changes can be studied by means of (a) ***Absolute historigrams*** which are obtained by representing the series as it is or (b) by means of ***Index Historigrams*** which are obtained by calculating the Index numbers of each series on the basis of figures of the year or the average of the figures a graph of years chosen as the base. In the case of absolute historigrams of two or more variables different scales are required to be taken in cases where it is not possible to work with one scale. In such cases great care and caution is necessary in choosing the different scales. The easier method in such a case is to resort to Index historigrams to avoid any complication. This will show proportionate changes in each case.

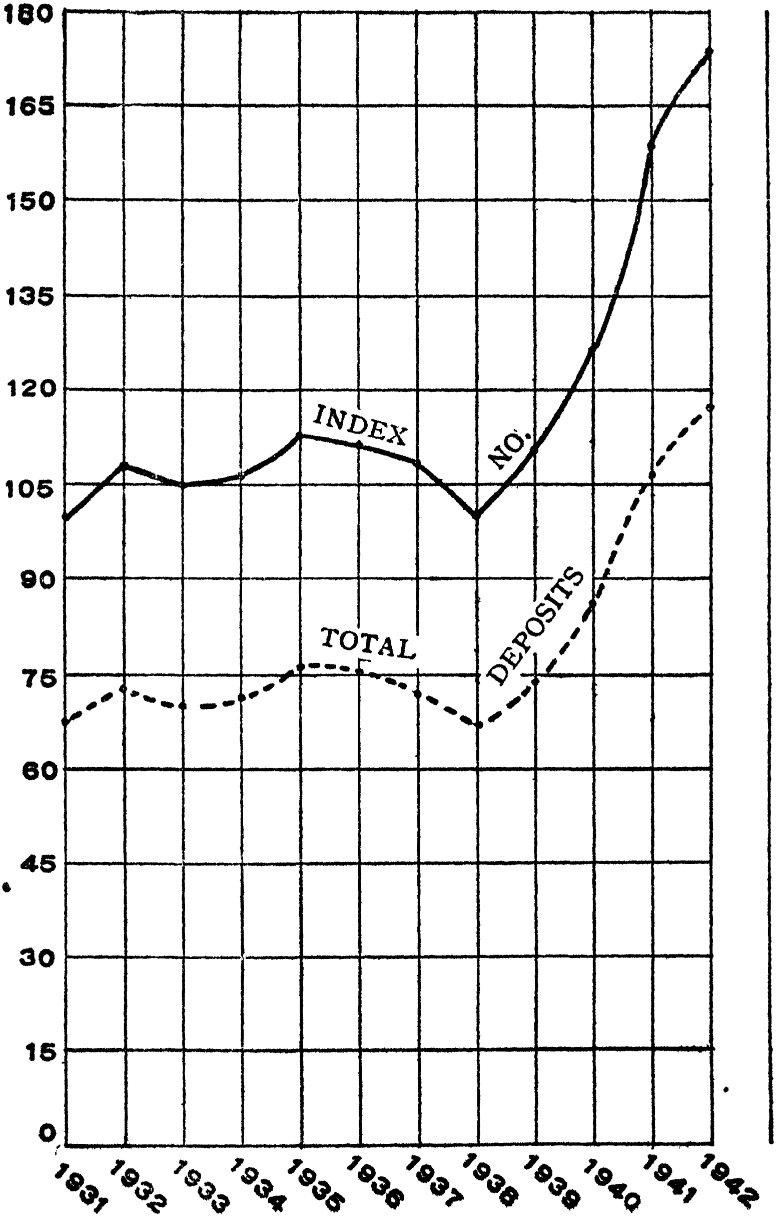
The following example will illustrate the graphical representation of one or more variables in both the forms.

Table 5.

Total Deposits of Exchange Banks secured in India.

Year	Rs (crores).	Index no.
1931	67.5	100
1932	73.1	108
1933	70.8	105
1934	71.3	106
1935	76.2	113
1936	75.2	111
1937	73.2	108
1938	67.2	100
1939	74.2	110
1940	85.3	126
1941	106.7	158
1942	116.9	173

Total deposit of Exchange Banks



Table—6. Production and export of Indian coffee
(in thousands of cwts.)

Years.	Production		Export	
	Cwts.	Index Nos.	Cwts.	Index nos.
1	2	3	4	5
1928	318	100	262	100
1929	248	78	138	53
1930	352	111	279	106
1931	294	92	208	79
1932	300	94	162	62
1933	295	92	169	65
1934	309	97	187	71
1935	293	92	148	56
1936	368	116	230	90
1937	304	96	162	62
1938	299	94	146	56
1939	358	113	207	79

In Fig. 5. graphs of absolute as well as of index numbers have been shown by different curves by taking different scales. In case of absolute historigrams correct idea of the proportionate rise and fall cannot be obtained which is easy to have from the Index number historigrams. This can be made further easy by drawing Logarithmic or Ratio-scale graphs.

Logarithmic Or Ratio-Scale Graphs

In the construction *of Natural or Ordinary Scale* **graphs**

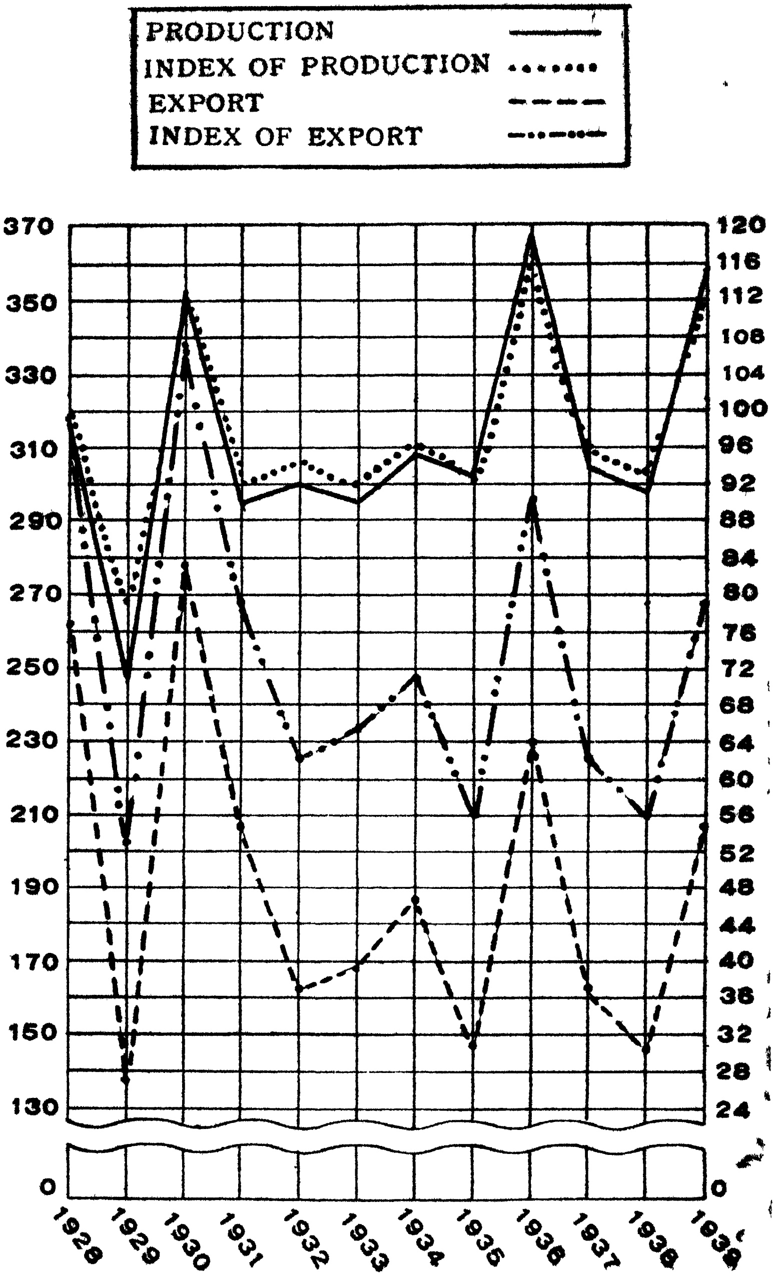

PRODUCTION
INDEX OF PRODUCTION
EXPORT
INDEX OF EXPORT
370
350
330
310
290
270
250
230
210
190
170
150
130
0
120
116
112
108
104
100
96
92
88
84
80
76
72
68
64
60
56
52
48
44
40
36
32
28
24
0
1928
1929
1930
1931
1932
1933
1934
1935
1936
1937
1938
1939

Fig. 5.

equal distance, represent equal absolute movements but for studying the *relative* movement *Logarithmic or Ratio-Scale* graphs are far more suitable, as therein, equal distances represent equal *proportionate* movements and not equal absolute movements as is done in the case of Natural scale graph as will be seen from Fig. 5. In this, the rise or fall of production is to be seen from the basic figure of 318, while that in the case of export it is to be compared with the basic figure of 262. However, the vertical distance of rise or fall are measured by the same units and so no proper indication of relative changes can be obtained.

A *Ratio-Scale or Logarithmic* graph can be drawn in one of the two ways ;—

(1) The logarithms of the variables are obtained from Logarithmic tables (*vide* Table no. 7) and they are then plotted on a natural scale graph paper as shown in Fig. 7. This method is quite simple in the case of persons who understand logarithms, but to an average person this is not easy and so the other method of plotting the variables on the semi-logarithmic graph paper will be more suitable.

(2) In the second method, the variables are not converted into logarithms but they are plotted not on the ordinary graph paper *but* on a ratio-scale ruled paper, (Fig. 8) in which vertical distances are marked on a logarithmic scale. This chart is known as 'Logarithmic Chart' though the correct designation is 'Semi-logarithmic chart' for the simple reason that the ratio-scale ruling is done only vertically and not horizontally as well. If the logarithmic ruling is done vertically and horizontally then only it can be really called 'Logarithmic'. In practice, however, it is customary to name the latter type 'Double Logarithmic' and the former as 'Logarithmic'.

Table 7.—**Indian cotton crop.**

Years.	Estimated yield in thousands of bales.		Estimated value (in lakhs.)	
	bales.	log.	Rs.	log.
1	2	3	4	5
1934–35	4857	3.6863	5158	3.7125
1935–36	5867	3.7684	6357	3.8033
1936–37	6234	3.7948	6961	3.8427
1937–38	5722	3.7576	4360	3.6395
1938–39	5051	3.7034	4331	3.6366
1939–40	4909	3.6910	5407	3.7330
1940–41	6080	3.7839	4906	3.6907
1941–42	6223	3.7940	4975	3.6968
1942–43	4702	3.6723	6700	3.8261

In Fig. 6, curves representing the estimated yield in bales and in money value have been drawn but the same data has been shown in Fig. 7, by curves based on logarithms of the same figures while in the third diagram, Fig. 8, curves have been drawn on a semi-logarithmic paper. In the first figure, equal distances are representing equal absolute movements, while in the last one they show proportionate movements which is a matter of vital importance for making comparisons.

Advantages and Disadvantages:—In the Ratio–Scale graph, as explained already, equal vertical distances represent equal proportionational changes and so for comparing the rise or fall in the case of two or more curves, this method is found to be very useful. In the natural scale graph, if the incomes of two persons getting Rs. 200 and Rs. 400 respectively rise by Rs. 100 in each case, the curves will show equal rise in both the cases. In fact, in the first case the rise is 50% while in the other, it is much less, being only

Fig. 6—Indian cotton crop:

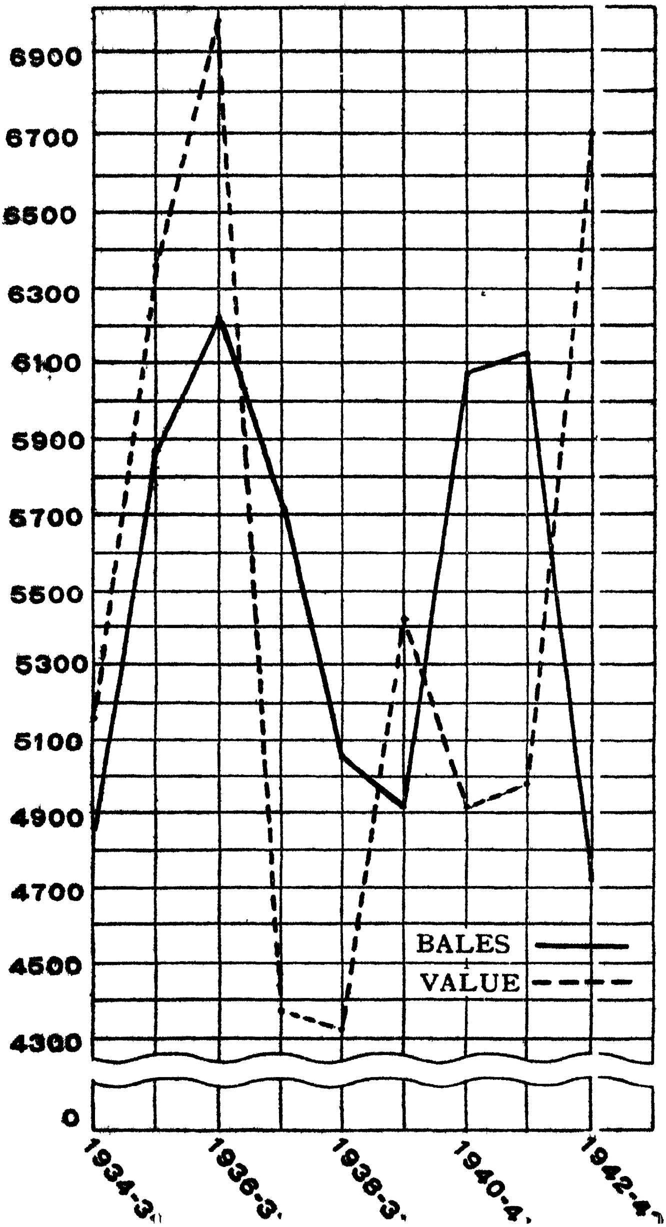
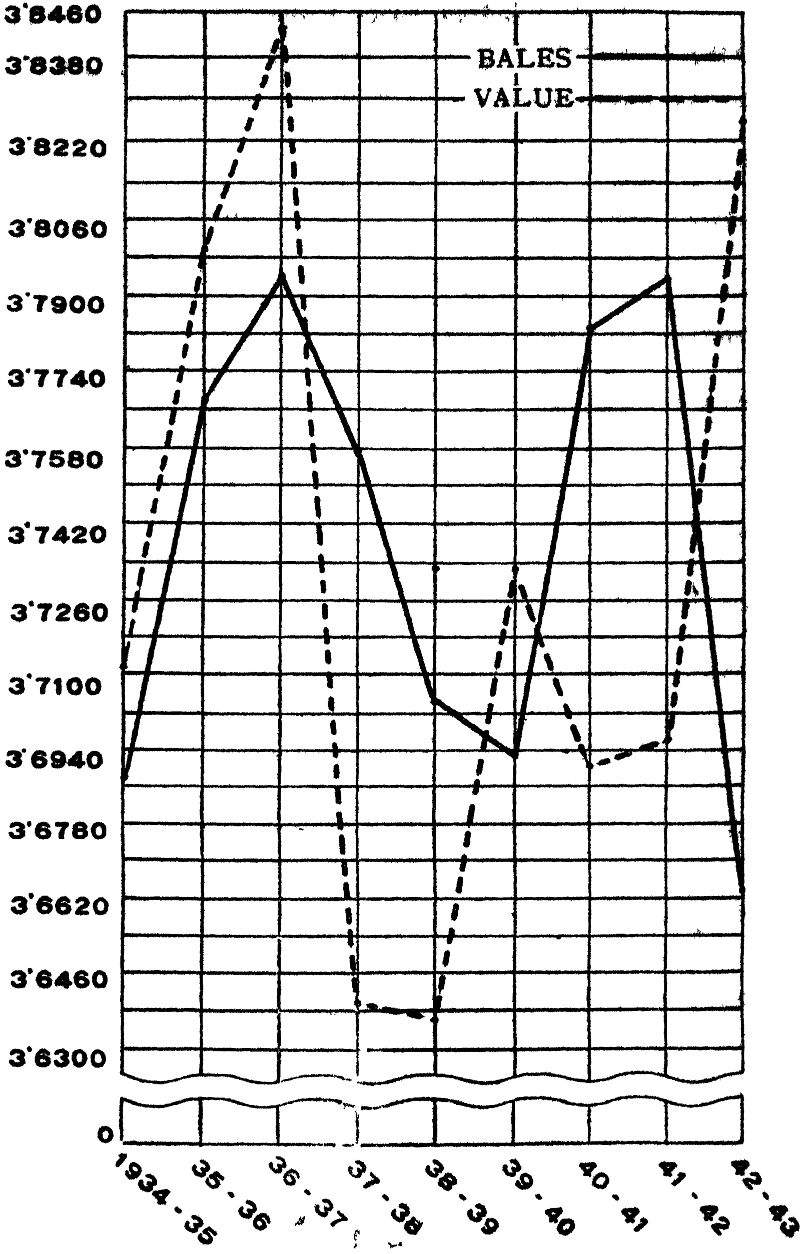


Fig. 7.—Indian cotton crop.



Indian cotton Crops.

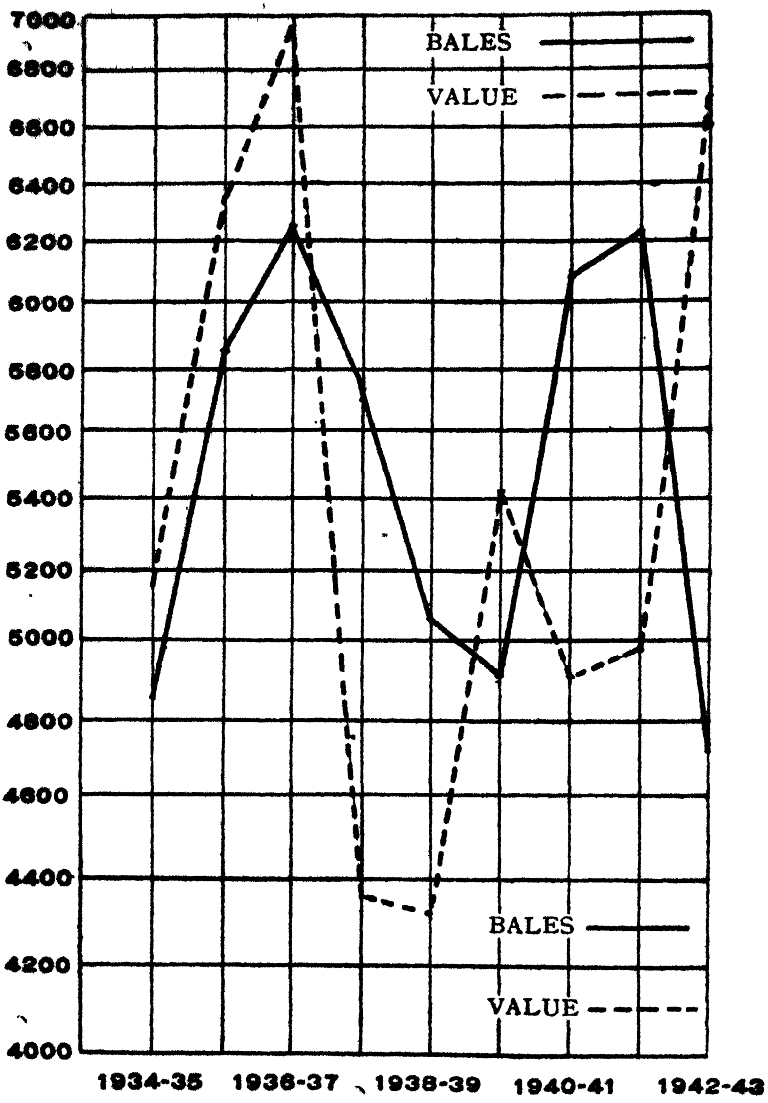


Fig. 8 Ratio scale graph.

25%. This can be easily distinguished by the use of ratio-scale graph.

If the variables show a great range of variation in the case of Natural Scale Graph, the ordinate of Y will have to be considerably lengthened, which is usually not feasible on the normal size of the paper. If this is not possible, the value in units per division of the scale will have to be increased, but this will mean that the rise and fall of the curve will be noticed with difficulty. In the ratio-scale chart, this difficulty will not arise, as equal distances will rise in equal proportion, as will be seen from the scale in Fig. 8. This will result in economy of space, besides being able to show a wide range of variation in the variables.

As the same vertical difference always shows the same percentage variation, diagrams can be drawn anywhere on the chart and thus shifting of the scale is easy.

On the logarithmic scale chart, constant rate of increase or decrease can be easily noticed and this property is of great use in extrapolation, particularly in the business life where forecasts have to be made on the assumption that the rise or fall will be at the same rate.

In ratio-scale chart, both the absolute values as well as proportionate changes can be easily marked. The proportionate change will be visually seen, while the absolute values can be ascertained by referring to the scale of the ordinate. For comparing two series of very small and very large quantities, this scale chart is very convenient, as both the curves can be easily represented.

In the ratio-scale, however, negative quantity cannot be represented which is so easy to do in the case of a natural scale graph. Again, there is no zero line in this and the comparisons of absolute sizes are not possible.

Comparison of the two types of graphs.—If we were to compare the Arithmetic or Natural scale graph with the Logarithmic or Ratio-scale graph, it will be seen that in the former chart, zero as well as negative values can be shown, which is not possible in the latter. In the case of variables having wide range of values, the ratio-scale graph is far more suitable than the other. If the aggregates as well as the constituents are to be shown, the ratio-scale is not convenient but in this, the curve can be drawn at any point without affecting the properties, which is not possible in the case of the natural scale graph. The chief point of difference is that in the natural scale, equal vertical distance will represent equal absolute distance, while in the other, it will show equal proportionate change.

The ratio-scale graph in spite of some of the advantages in its favour, does not usually appeal to an average reader which may be probably due to his lack of familiarity with logarithms.

Other types of statistical charts :—We have so far discussed the construction of the various graphic charts pertaining to the frequency and time series as well as the ratio-scale charts. We shall now describe other types of less importance not discussed so far.

Galton's graphic method of locating median :—As already pointed out in the chapter on 'Averages' median can be located by the graphic method as suggested by Galton. On the horizontal line, class intervals are marked and the vertical line is used for marking the frequency. It is more convenient if the observations are not grouped into classes. If there are more observations for a particular class, more dots are put in the vertical line as shown in Fig. 9. It should be carefully noted that at the time of plotting the points, every preceding measurement is taken as the base for the succeeding one.

After all the points are plotted, a curve is drawn connecting every set of successive dots on the vertical line. If there is only one dot, the curve will be made to pass through it. If there are two or more dots but consisting of an even number, the curve will be drawn passing mid-way between the two middle dots. Where the dots are odd in number, the line is drawn through the central dot. After the continuous curve is drawn, a horizontal line parallel to the base will be drawn first from the middle of the vertical line measured from the base to the last point in the scale. A perpendicular will then be drawn to the base from the point where the curve is intersected by the parallel line. The distance from the origin to the foot of this perpendicular measured on the base line will give the value of the median.

Table 8—**Age of 35 students in a school.**

Roll no.	Age	Roll no.	Age	Roll no.	Age
1	7	13	12	25	15
2	8	14	12	26	15
3	8	15	12	27	15
4	9	16	12	28	16
5	9	17	13	29	16
6	9	18	13	30	16
7	10	19	13	31	17
8	11	20	13	32	17
9	11	21	13	33	18
10	11	22	14	34	18
11	12	23	15	35	19
12	12	24	15		

Locating a median.

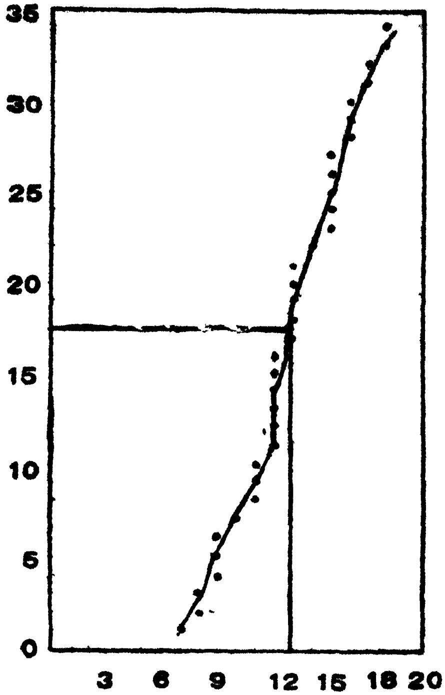


Fig. 9. Galton graph.

The Zee Chart or **'Z'** chart is a multiple curve chart and is so called because it consists of three curves resembling the letter **'Z'**. These three curves are (1) the curve of the original data, (2) the cumulative curve and (3) the moving total curve. Different scales will have to be taken to suit the plotting of the variables to suit the size of the diagram. On the top will be the curve of the moving total while the curve of the original data will be at the bottom and the cumulative curve will connect the two to form the letter **'Z'** as shown in Fig. 10.

Table 9.—**Sales** (in thousands of rupees).

	Jan.	Feb.	Mar.	April	May	June	July	Aug.	Sept.	Oct.	Nov.	Dec.
Sales in 1944 ...	150	160	145	170	175	190	185	135	145	165	180	190
Cumulative ...	150	310	455	625	800	990	1175	1310	1455	1620	1800	1990
Moving Annual Total. ...	—	—	—	—	—	—	—	—	—	—	—	—
Sales in 1945 ...	130	150	170	160	140	170	190	130	160	150	190	200
Cumulative. ...	130	280	450	610	750	920	1110	1240	1400	1550	1740	1940
Moving Annual Total ...	1970	1960	1985	1975	1940	1920	1925	1920	1935	1920	1930	1940
Sales in 1946 ...	190	180	200	210	190	190	210	170	180	180	200	220
Cumulative ...	160	340	540	750	940	1130	1340	1510	1690	1870	2070	2290
Moving Annual Total ...	1970	2000	2030	2080	2130	2150	2170	2310	2230	2260	2270	2290

The moving annual total is obtained for Jan. 1945 by dropping Jan. total of 1944 and adding that of Jan. 1945 and for Feb. 1945 by dropping Feb. 1944 and taking Feb. 1945 and so on continuously for all other succeeding totals.

Sales in 1949

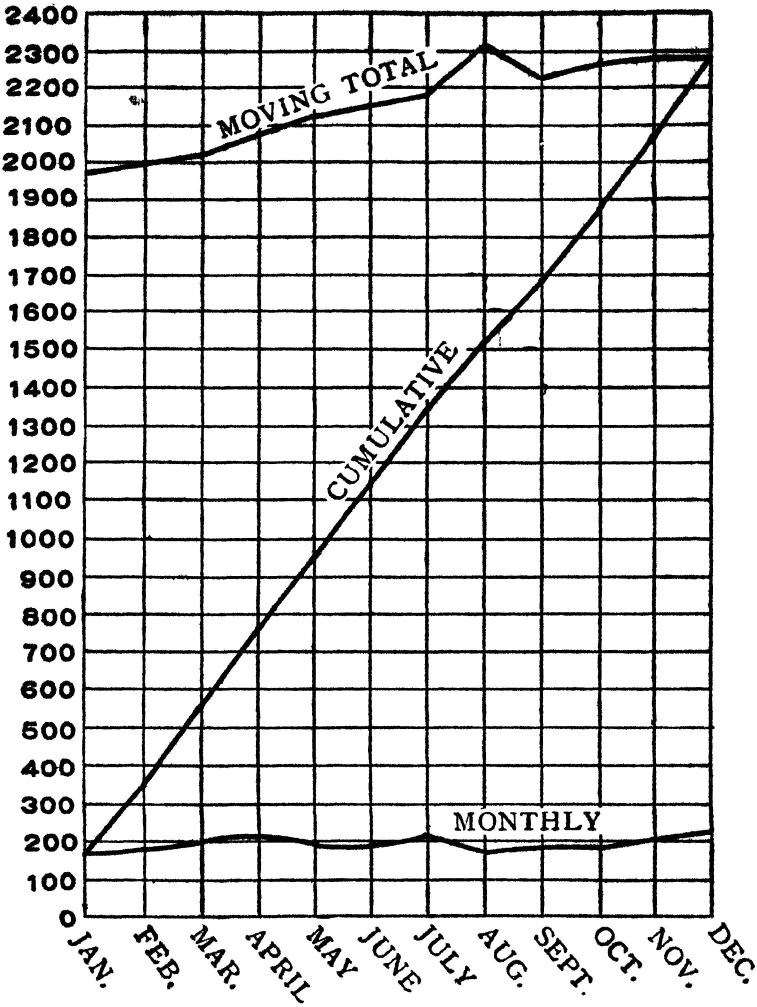


Fig. 10. Zee chart

Zone curve:—A zone curve is obtained by plotting curves for the minima and the maxima and then filling in the intervening space between the two curves with shading or vertical lines.

Table 10.—**Prices of silver in the Bombay Market**

Year	Highest Rs.	Lowest Rs.
1932–33	60– 0–0	48–14–0
1933–34	59–14–0	52–15–0
1934–35	69– 0–0	50– 7–0
1935–36	87– 0–0	46–12–0
1936–37	56– 2–0	47– 0–0
1937–38	55–15–6	46– 1–6
1938–39	53– 1–0	48– 2–0
1939–40	66– 4–0	44– 7–0
1940–41	64–13–0	62–11–0

Prices of silver.

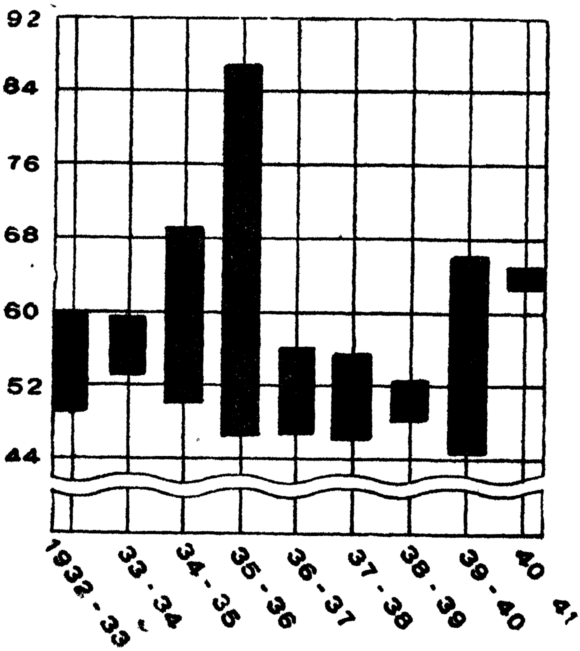


Fig. 11. Zone curve.

Band or Belt curve:—In the band or belt curve chart, the component parts of the whole are plotted one above the other and the different bands or belts are distinguished by different shadings or colours.

Table—11. **Imports of principal commodities in India.**
(in crores of rupees)

	1938–39	1939–40	1940–41	1941–42	1942–43
I. Food, drink etc.	24.0	35.3	23.8	27.9	7.6
II. Raw materials etc.	33.2	36.1	42.1	50.0	52.0
III. Articles–mainly manufactured.	92.8	91.8	89.5	93.5	49.5
IV. Miscellaneous.	2.4	2.1	1.6	1.7	1.3
	152.4	165.3	157.0	173.1	110.4

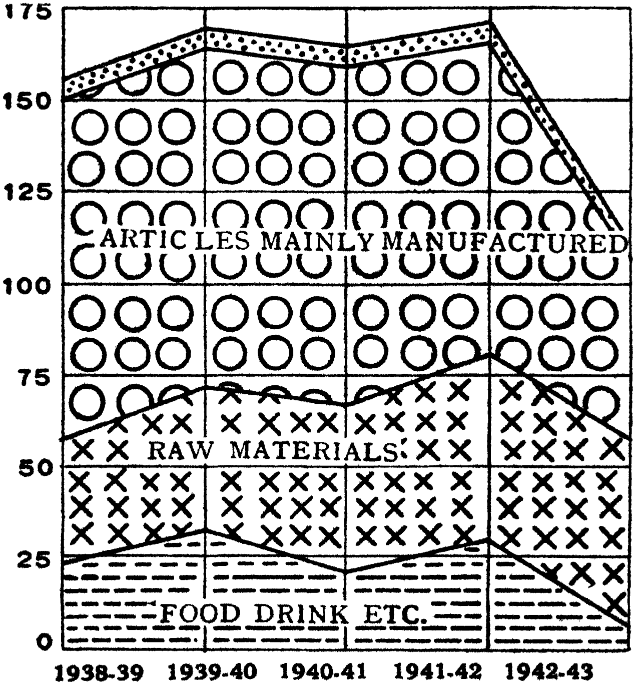


Fig. 12. Band curve.

If so desired, the data can be shown in percentage form, but in that case, the whole chart will represent only 100 and the percentages of the component parts will have to be taken.

'Surface' and 'Strata' charts:— When the space below the curve is filled in with some kind of colour or shading, the presentation is often called a 'Surface chart' and if this chart is further sub-divided into layers or strata in proportion to its components it is known as a 'Strata chart'. It is in fact the band curve chart as shown in the above figure.

The Lorenz curve:—It is a cumulative frequency curve designed to show the deviations in the distribution and so its construction has been explained in the chapter on 'Dispersion.'

Exercises.

1. 'The wandering of a line is more powerful in its effect on the mind than a tabulated statement.' Discuss this statement by giving suitable examples.
2. Give the important rules to be followed in constructing a suitable graph.
3. Differentiate between
   (*a*) a histogram and a frequency polygon, and
   (*b*) a histogram and a historigram.
4. Outline the method of drawing a smoothed frequency curve.
5. How are the changes over a period of time to be studied graphically?
6. What is a ratio-scale graph? State how it differs from the natural scale graph and discuss the merits and demerits of each type.
7. Outline the procedure for constructing a logarithmic curve.
8. Is it possible to construct a logarithmic curve without ratio-scale ruled graph paper? If so, how can it be done?
9. Write short notes on:—
   Independent variable, False base line, Logarithmic chart, Zee-chart and Band curve.
10. How will you locate a median by means of Galton's graph? Illustrate your answer by an example.

11. Show graphically the normal monthly maximum temperature in shade and also the normal monthly rainfall at the following stations :—

	Jan.	Feb.	March	April	May	June	July	Aug.	Sept.	Oct.	Nov.	Dec.
Temperature												
Simla…………	46.4	46.8	55.2	64.6	72.1	73.1	68.9	66.7	65.8	62.7	56.0	49.8
Bombay………	82.9	82.9	85.8	88.5	90.8	88.3	85.4	84.9	85.3	88.7	89.2	86.4
Calcutta………	77.5	82.3	91.0	95.5	94.6	91.3	88.6	87.8	88.2	87.4	82.2	77.0
Rainfall												
Simla…………	2.7	3.1	2.7	1.9	2.9	7.1	16.9	17.3	6.2	1.1	0.5	1.1
Bombay………	0.1	0.1	0.1	0.1	0.8	18.3	24.3	13.8	10.5	2.2	0.4	0.1
Calcutta………	0.8	1.1	1.4	1.9	5.8	11.9	12.5	12.7	9.9	4.2	0.7	0.2

12. **Represent the following graphically :—**
Export of Indian raw cotton in thousands of tons

	Total exports	**Exports to Japan**
1926–27	569	221
1927–28	479	287
1928–29	663	293
1929–30	727	329
1930–31	701	301
1931–32	423	193
1932–33	365	194
1933–34	504	197
1934–35	623	367
1935–36	606	314

(1940–Mar.)

13. The following table gives the age distribution of widowed females in India. Draw a graph showing the number of widows younger than any given age and from the graph find the median of widows :—

Age group	Widows (in '000)	Age group	Widows (in '000)
0–5	31	40–45	3081
5–10	105	45–50	3399
10–15	185	50–55	3110
15–20	533	55–60	2[illegible]99
20–25	877	60–65	2487
25–30	1580	65–70	1257
30–35	2000	70 and over	1958
35–40	2848		

(1944–Mar.)

14. The following table gives the proportion of married women in 1910 and in 1920 from women of every age. Show graphically that the increase was most marked for the women of younger years.

Percentage of married women

Age	1910	1920
18	17.0	19.2
20	36.2	38.4
22	50.7	52.9
24	62.0	64.2
25	65.7	67.8

(1941–Oct.)

15. The following table gives the manufacture of gunny bags in the millions. Represent this data graphically on natural as well as logarithmic Scales.

1919–20	343	1928–29	498
1920–21	534	1929–30	522
1921–22	387	1930–31	434
1922–23	344	1931–32	389
1923–24	414	1932–33	415
1924–25	425	1933–34	402
1925–26	425	1934–35	423
1926–27	449	1935–36	459
1927–28	463		

16. The following table gives index numbers since 1925 for Calcutta, Bombay and Karachi (Base July 1914=100.) Show these by means of a suitable graph.

Year	Calcutta	Bombay	Karachi
1925	159	163	151
1926	148	149	140
1927	148	147	137
1928	145	146	137
1929	141	145	133
1930	116	126	108
1931	96	109	95
1932	91	109	99
1933	87	98	97
1934	89	95	96
1935	91	99	99
1936	91	96	102
1937	102	106	108

17. The following table shows the number of persons assessed to income-tax in a province of British India, classified according to grades of income. Find *graphically* the

income level reached by 30 per cent and 95 per cent of all persons assessed to income-tax.

Grades of income	Number	Grades of income	Number
Rs. 1000 to 1499	3,015	Rs. 12500 to 14999	151
„ 1500 „ 1999	2,185	„ 15000 „ 19999	144
„ 2000 „ 2499	1,034	„ 20000 „ 24999	49
„ 2500 „ 2999	498	„ 25000 „ 29999	31
„ 3000 „ 3499	376	„ 30000 „ 39999	25
„ 3500 „ 4999	791	„ 40000 „ 49999	7
„ 5000 „ 7499	786	„ 50000 „ 99999	5
„ 7500 „ 9999	408		9,727
„ 10,000 „ 12499	222		

(1941 Mar.)

18. Draw a graph to compare the amounts of cheques cleared annually in three commercial towns as shown in the following table :

(In crores of rupees)

Years	Calcutta	Bombay	Madras
1927	1024	328	56
1928	1088	543	65
1929	998	800	59
1930	893	712	52
1931	756	640	45
1932	747	646	47
1933	824	646	52
1934	864	683	58
1935	939	750	63
1936	899	721	84
1937	993	837	109

19. Represent graphically the exports and imports of India from the following table on the Natural as well as on the Ratio Scale.

In crores of rupees

	Exports	Imports
1927–28	319	250
1928–29	339	263
1929–30	345	258
1930–31	308	206
1931–32	263	176
1932–33	239	203
1933–34	275	182
1934–35	280	210
1935–36	282	216
1936–37	343	199

20. Draw a graph to compare the movements in the following Index Numbers and comment:

Month		Economic Adviser's Index of Wholesale Prices	Bombay City Cost of Living Index
Sept.	1939	113	106
March	1940	128	110
Sept.	1940	111	112
March	1941	119	119
Sept.	1941	142	129
March	1942	146	137
Sept.	1942	167	170
March	1943	220	208
Sept.	1943	236	245
March	1944	232	216
Sept.	1944	242	239
March	1945	248	225

(1946-Mar.)

CHAPTER IX

HISTORICAL OR TIME SERIES

A statistical series is an orderly set of numerical items expressing relationship between two variables and the one having a numerical record of changes for a variable during a number of successive intervals of time is known as a ***Time or Historical Series.*** The graph obtained by plotting such a series is called 'Historigram', the various types of which have been already described in the chapter on 'Graphs'.

Of late, Time series are attaining considerable importance particularly in economic and commercial matters. One of the major problems in the modern business life is that of estimating the nature of future changes. As usually these changes are not very irregular, but are ordinarly based on somewhat regular movements such as evolutionary, periodic and cyclic, it becomes necessary to analyse the data over a period of time with a view to obtaining some clue to the future movements. This is achieved by a careful and analytical study of historical series which is mainly the purpose of having a separate chapter on this topic.

There are two types of Historical or Time series. ***Simple Historical Series*** are those in which amounts or frequencies relate to instants or intervals of time while in ***Cumulated*** Historical Series amounts or frequencies are added at successive instants or intervals of time. If sales of a shop for January, February, March and April are Rs. 8,700, Rs. 9,300, Rs. 10,200 and Rs. 9900 respectively, they will constitute a simple historical series of the sales of that shop. In cumulated historical series the amounts will not relate to instants or intervals of time but to their totals at successive intervals and so the series formed in this case will be as under :—

Cumulated historical series of sales

January		Rs. 8700
February	(8700+9300)	Rs. 18000
March	(8700+9300+10200)	Rs. 28200
April	(8700+9300+10200+9900)	Rs. 38100

However, it is not possible to form such cumulated historical series in all cases. For instance, there can be a cumulated series of rainfall but not of temperatures or of market rates. It will be simply absurd to add temparatures of the various months and to form a series as this will serve no useful purpose.

Classification :—From a tabular representation it is difficult to discover the various trends that exist in most of the time series. But if represented graphically they often exhibit more or less regular fluctuations at intervals and appear to follow some direction or trend. Such tendency is particularly visible in economic and commercial data regarding sales, prices, interest and bank rates, unemployment etc. No doubt, these movements or fluctuations are usually the result of a combination of various influences such as evolutionary, cyclical, seasonal and the like, It is very rare to find a time series influenced by a single element only and due to this complexity it is difficult to assess accurately influences exerted by different elements involved therein. Broadly speaking, these movements can be grouped into the following four important types, one or more of which are responsible for movements of time series.

1. Basic or Long–time Trend
2. Undulatory or Cyclical Movement
3. Periodic or Seasonal Variation
4. Irregular Fluctuations

(1) **Basic or Long-time Trend** :—There is a general tendency in nature of things to gradually increase or decrease over a period of several years. For instance, business concerns or factories normally expand gradually in the beginning in spite of temporary adverses. Population as well as production normally shows growth. Equally, in some cases like average length of life, rainfall etc. gradual decline is also visible. This underlying persistent tendency to grow or to decline over a period of years is known as *Basic or Long-time Trend.* It is also called *Primary Movement, Trend, Secular Trend or Evolutionary Trend.* All basic trends are not of the same nature and do not all the while work in the same direction. In one period this trend may be influenced by

the growth-factor while in the other the decline factor might set in. As said already, this upward or downward tendency can scarcely be the only factor influencing the time series, though it is equally true that this is the only factor that will persist in its influence for a considerably long period.

(2) **Undulatory or Cyclical Movements** :—It has been observed mainly in economic and commercial data that in addition to the secular trend a fairly regular upward and downward tendency is visible at intervals ranging normally from seven to eleven years. This period is evidently much less than the one covered by secular trend but is at the same time far longer than that of seasonal variations. This cyclical movement lacks in regular periodicity, but moves sporadically in response to causes which develop intermittently out of complex combinations of economic and other circumstances. If the data is represented graphically, business cycles, wave-like in character, are apparent. The crests represent the boom years while the troughs show the years of depression. Because of seasonal influences these waves appear to change their form to some extent. These cyclical changes, in short, are due to swings in business through periods of expansion, liquidation, depression and recovery covering in all seven to eleven years. This movement which shows tendency of history to repeat itself is some times designated as *Secondary movement.*

(3) **Periodic or Seasonal Variation**—Along with the growth or decline factor and cyclical movements already explained, a definite recurrence of fluctuations frequently of a regular periodic type is visible daily, weekly, monthly or seasonally depending upon the nature of the series. The change is known as *Periodic, Seasonal or Short-time Variation.* It is usual to find prices going down in the harvesting season and slowly rising thereafter, particularly in the case of agricultural products. In summer, price of vegetables soar up and are found to ease sometime after the rainy season. On week days in busy towns particularly near about 10 A. M. there is a great rush in local trains while on Sundays and holidays it is not so. There is a great demand for woollen cloth near about winter every year while umbrellas sell well in the rainy season. Seasonal variations,

are also visible in India in case of labour. If the rainy season is satisfactory many workers go to their villages and flock in towns for employment when the agricultural season is over. These variations are usually determined by measuring the actual occurrences over a certain period, assuming, of course, that the season or period is normal. Such changes are also classed as *Tertiary.*

(4) **Irregular Fluctuations** :—In this group are included all movements residual, episodic as well as fortuitous or accidental type. There are movements due to specific causes ordinarily reflected in sharp pronounced breaks in the record of variables with no apparent tendency for recurrence at stated intervals. For instance, due to strikes, production figures are seriously affected. Fortutious or accidental variations which are due to irregular disturbances such as wars, floods, earthquakes upset the normal activities. These movements or fluctuations cannot be traced to the steady influence of a secular trend or to rhythmical influences of business cycles or to seasonal or periodical variations. These movements are very complex in character and are often not separable from the business cycle movements.

Measuring time changes:—As already remarked, in historical or time data are found types of variations exerting a combined influence and it becomes an intricate problem to segregate the particular results and causes in which we are interested. If we are to make an analytical study of cloth prices extending over a number of years it will be noticed that apart from the normal tendency of growth or decline, it has been influenced by cyclical as well as seasonal movements and casually by irregular fluctuations due to episodic or accidental factors such as strikes, war, diseases, etc. A statistician is less fortunate than his brother chemist or biologist who can carry on his experiments under controlled conditions. In economic and commercial studies the data is to be taken as found and variations due to causes in which we are not interested are to be assessed and excluded thus giving us a residue which is chiefly the result of the cause under investigation. Different methods are used for different types of variations; the most important of them as applied in the various causes have been given in the following pages.

It must, however, be noted that before any method is adopted, it is necessary to guard against taking isolated extreme items as well as choosing a wrong and abnormal period. Again, to obtain a reliable result the period should comprise of a sufficiently long time to bring out properly the variations which are to be studied.

Measurement of Long-time trend :—This tendency of growth of decrease extending over a long period continues in spite of cyclical, seasonal or irregular fluctuations and is usually based on arithmetic, geometric or S-curve progression. In the case of arithmetic progression the increase or decrease is approximately by the *same amount* and if graphically presented it takes the form of a straight line. In geometric progression the rise or fall is approximately by the *same rate* or precentage in each period and if a logarithmic curve is plotted a straight line will be obtained. In S-curve progression a *logistic* curve is formed in which a slow rise in the beginning is followed by the rapid growth till it reaches a point of maturity when the rise again becomes rather slow.

The 'S-curve' :—This curve is usually obtained by plotting the figures of the growth of a new industrial concern. At the outset the progress particularly on per capita basis is rather slow and later on it increases till it reaches the market saturation point when it again becomes slow. A cumulative curve of this phenomenon, if drawn, naturally assumes the form of 'S' and hence this nomenclature.

Methods adopted :—In order to ascertain the long-time evolutionary, or secular trend it is necessary to exclude short-time as well as irregular oscillations from the time series under consideration. For this purpose the principal methods made use of are following :—

(a) Free-hand smoothing
(b) A Series of averages
(c) The moving average
(d) The Method of Least-squares

(a) **Free-hand smoothing** :—When a curve is drawn its course is rarely smooth as there are peaks and troughs due to fluctuations. For eliminating these oscillations the most easy device is of smoothing them by drawing a free-hand curve which will indicate the general tendency or trend. This is, no doubt, a dangerous device if left to an inexperienced hand as different persons are likely to draw different curves based on the same data. To guard against this it is better to draw three free-hand curves, two of which should represent the extreme variations in either direction and the third should be as a sort of compromise between the two. This will naturally be the most appropriate curve as the two extreme curves will serve as guide in its construction.

Table. 1 :—**Profits of Coal Industry** (in lakhs of rupees)

Year	Profits	Year	Profits
1928	80	1936	53
1929	79	1937	61
1930	98	1938	123
1931	77	1939	119
1932	64	1940	120
1933	61	1941	98
1934	50	1942	94
1935	54	1943	113

No. doubt this method is very practical as well as easy but it is defective as, given the same data, no two persons will be able to draw identical smoothed curves even if they have no ulterior motive to do otherwise. From the point of accuracy, therefore, this method is found to be deficient and also not very reliable.

(b) **A series of averages** :—To remove the uncertain character of a free-hand curve this method can be used with advantage. In this the whole series is divided into suitable sections and averages are ascertained for such sections. These points are then plotted and a free-hand curve is drawn passing through these. As these points are to be the same in all cases, this smoothed curve drawn by several persons will not be very much different. In Table 1 giving profits of coal industry the data is for sixteen years. It can be conveniently divided into four sections and then the smoothed curve will be drawn through the four points of averages 83.5, 57.3, 89.0 and 106.5 which will be plotted on the middle ordinates of each section. The following figure is obtained by applying this method to the preceding data.

Profits of Coal Industry in India

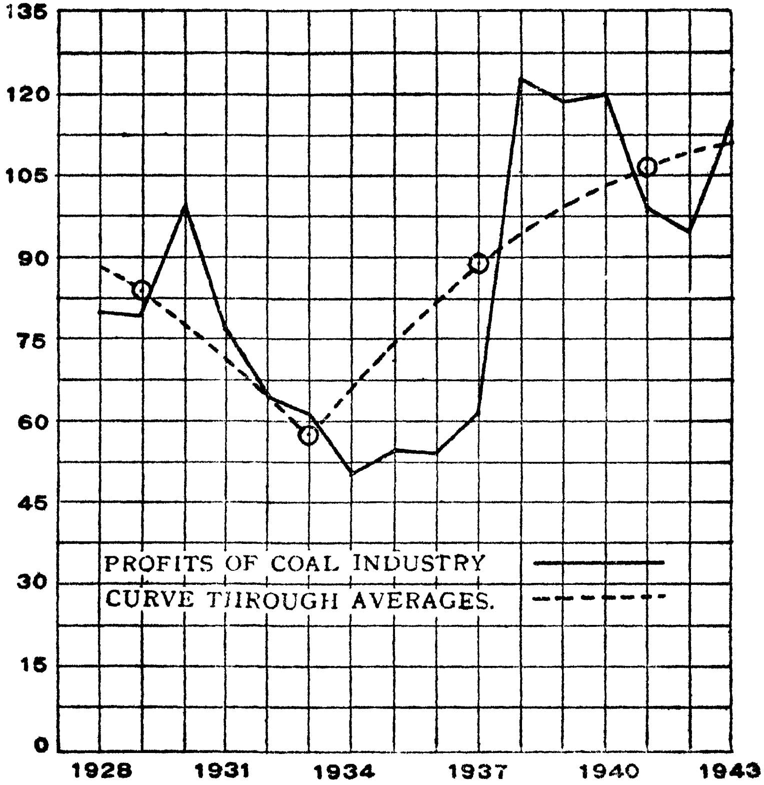


Fig. 1.

Method of Semi-averages:—Instead of dividing the whole data into several parts it is divided into two sections only. The smoothed curve is in this case drawn through the two averages plotted on the middle ordinate of each section. This method is comparatively simple but its result will be naturally less accurate.

(c) **The Moving Average** :—It is a form of arithmetic average and is calculated by replacing the first item taken for computing the average ascertained by the newly arising item for calculating the next average. For instance, if the triennial moving average is to be calcuated from a series of ten years arithmetic average will be computed for the first three years and will be entered against the second year. The next average will also be of three years but not of the first three. The average of the second, third and fourth years will be taken and written next i. e. against the third year. In this way the moving average of the whole series will be ascertained. The average of eighth, ninth and the tenth years will be written against the ninth year and there will be no average to write against the tenth year unless we are given the figure for the eleventh year. In the same way Quinquennial and Septennial year moving averages can be computed the only difference being that instead of three, five or seven year figures will be taken for computing the average. It is usual to find moving average computed for 3, 5, 7, and 10 years through it can be obtained for any number of years. The choice of the period will mainly depend upon the problem in hand. In the case of an even number of years the difficulty experienced is about the place of writing the average. In the case of an odd number it is written against the middle item. In the case of decennial moving average it cannot be written against the fifth or the sixth year as the nine-yearly and the eleven-yearly moving averages if calculated will be written there. To get over this difficulty the moving average is *centred* by taking the average of succeeding figures of moving averages and then writing it against the middle item the first year against which it will be written in this case will be the sixth year as is done in column (8) of table no.2. Some use the method of 'lagging the moving average,' i. e. of writing it against the previous year instead of writing in the middle of the two years but this method is not suitable particularly when the deviations have to be computed.

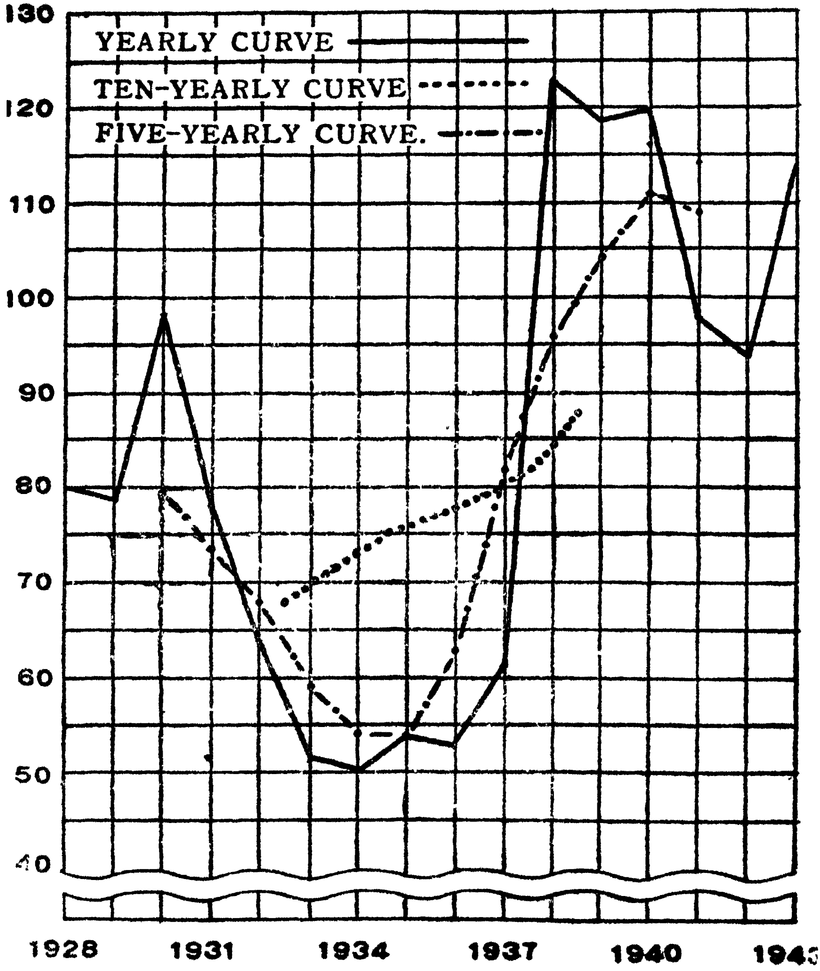
These averages should then be plotted and the smoothed curve thus obtained will give the trend of the data by disentangling the short-term oscillations. In the case of a moving average the disadvantage is that it cannot be kept quite up-to-date as the moving average figure for the last few years cannot be computed in the absence of figures for succeeding years. In the case of a series for a long period trend this cannot be considered as a very great handicap.

Table-2--**Profits of Coal Industry in India** (in lakhs of rupees)

Year (1)	Profits (2)	Sums in 5's (3)	Five yearly moving average (4)	Sums in 10's (5)	10 yearly moving average (6)	add in pairs (7)	Centered average (8)
1928	80	...	...				
				...	. .		
1929	79	...	...	...			
			...	...	...		
1930	98	398	79.6				
	...	...	...	...	...		
1931	77	369	73.8				
		...	...	...	...		
1932	64	340	68.0				
	...	...	...	667	66.7		
1933	51	296	59.2	...	...		
	...	...	...	710	71.0	137.7	68.9
1934	50	272	54.4	...	...		
	...	...	...	750	75.0	146.0	63.0
1935	54	269	53.8	...	...		
	...	...	...	772	77.2	152.2	76.1
1936	53	341	68.2	...	...		
	...	...	...	793	79.3	156.5	78.3
1937	61	410	82.0	...	...		
	...	...	...	823	82.3	161.6	80.8
1938	123	476	95.2	...	...		
	...	...	...	886	88.6	170.9	85.5
1939	119	521	104.2	...	...		
	...	...	...	..	...		
1940	120	554	110.8				
	...	...	...	...	...		
1941	98	545	109.0				
	. .	...	...	...	...		
1942	94	...	...				
	...	...	...	...	...		
1943	114	...	...				

In the above table both quinquennial as well as decennial moving averages have been given. In the case of the latter in column (6) is shown the 10-yearly moving average written between two years as it cannot be written against any particular year due to its being the average of an even number of years. In column (8) the centered average figures are given and are written against the different years.

Profits of coal industry in India



As will be clear from the graph showing the profits of coal industry for various years as well as its five and ten yearly curves that the trend is to a great extent visible. The averages plotted for getting a ten-yearly curve are from column (6). As the figures for 1944 and other years are not given, the moving average cannot be computed beyond 1941 and 1938 in the case of quinquennial and decennial moving averages. If these curves are drawn on semi-logarithmic paper the trend obtained will be more reliable.

If the figures available are for few years only it is not convenient to compute moving averages and so in this case ' Progressive Average ' can be made use of.

Progressive Average :—This is a form of average ascertained by adding different items and dividing the result by the number of items taken. A column is provided to write increase or decrease of the particular year's figure as compared to the progressive average as calculated at the end of that particular period.

Table 3.—**Sales of the India Stores Ltd.**

Year	Sales Rs.	Progressive average Rs.	Increase or decrease of years sale on average. Rs.
1935	37,000	—	—
1936	39,000	38,000	1000
1937	44,000	40,000	4000
1938	48,000	42,000	6000
1939	42,000	42,000	—
1940	36,000	41,000	—5000
1941	48,000	42,000	6000
1942	50,000	43,000	7000
1943	47,000	43,500	4000
1944	38,000	42,950	—4950

It will be seen from the above table that in 1939 there is no decrease or increase but in 1940 and 1944 there has been a decrease of Rs. 5000 and Rs. 4950 respectively. This has been shown by means of a minus sign but normally such figures are shown in red ink in statements.

This form of average is particularly useful during the early years of a concern as it is difficult to have moving averages for want of data extending over a number of years. This form of average suffers from the defect that it includes data for earlier years which loses to a great extent its importance as it gets old. For this reason it is necessary that it should be discontinued in favour of the moving average as soon as it is convenient to do so.

(*d*) **The Method of Least Squares:**—This method is a mathematical device and by this a line is drawn through plotted points in such a way that the sums of the squares of the deviations of the actual points above and below this '*line of best fit,*' are at a minimum. The determination of a straight line for this purpose by the method of least squares is illustrated by the following example.

Table 4—**Consumption of Cotton in British India**

(in lakhs of bales)

Year (a)	Consumption in lakhs of bales. (b)	Time deviation from the mid-year. (c)	Deviation of col. (c) squared. (d)	Product of cols. (b) and (c) (e)	Trend ordinates in lakhs of bales (f)
1941	42.5	–2	4	–85.0	44.44
1942	47.4	–1	1	–47.4	45.87
1943	49.0	0	0	0	47.30
1944	48.5	1	1	48.5	48.73
1945	49.1	2	4	98.2	50.16
Total Average	236.5 47.3		10	14.3	236.50

The arithmetic average of the items should be ascertained which would then be the middle point of the line representing the trend. The middle item of the series in point of time should be marked as 0 and from that mid-year deviations should be marked in column (c) those before the mid-point being considered minus. Values of items in column (b) should be multiplied by the deviations as shown in column (c) and the product entered in column (e). The total of this product should be divided by the sum of squares of the deviations as written in column (d). The result will give the average amount of increase in the trend from year to year. As already stated the arithmetic average will be written against the mid-year. To this should be added the amount of increase arrived at year after year for each succeeding year and for each preceding year it should be deducted. It will be seen that the total of column (f) in which these figures are written would be equal to the total of column (b). The values in column (f) should then be plotted to have the '*line of best fit.*'

Measurement of Other Fluctuations:—After eliminating the variation of the long time trend from the original data by one of the methods already suggested and by extracting the erratic or irregular movements by taking a data covering a slightly longer period or by taking data having a wider scope the influence left on the data can be ascribed to regular short-time fluctuations due to cyclical and seasonal variations. The cyclical influence then can be ascertained by eliminating the seasonal variation method which is discussed hereafter.

It has been observed that business cycles normally covering seven to eleven years recur constantly and possess four well-defined phases or periods: prosperity, decline, depression and improvement. In the prosperity period the public is optimistic and as the prices are high, easy profits are made and in general there is a boom. Expansion of business then sets in to the extent of overdevelopment. The demand for money increase and business men offer concessions in price to secure ready cash. Buyers try to wait for still more lowering of prices. Interest rates go up and prices begin to drop down. All this leads to a decline of business. Very often a panic sets in because of a heavy decline in business leading to severe financial strain and abnormal liquidation of holdings. A period of depression follows with general pessimism in trade and industry. If the depression is of a lighter type it is referred to as *recession*. The situation then gradually improves because of rigid economy and reorganisation of money and credit due to a general cleaning up. This improvement slowly regains the old prosperity and then again there is a similar cycle. In short, this business cycle is a phenomenon in which there is a recurrence of good and bad times.

The weather and climatic changes in different parts of the year influence directly or indirectly the economic and commercial activities. In certain seasons the productive activity particularly in agriculture increases while in certain parts of the year decrease in selling activity is easily noticeable. The recurrence of seasonal variation is almost regular and does command attention of statisticians and if observed carefully one would find peaks and troughs in curves at regular intervals. Two features appear more outstanding in this type. Firstly, there exists usually a certain *proportional* deviation from the average rather than a definite amount and secondly the changes in this proportional deviation are ordinarily slow.

Methods Adopted :—Before any method is adopted it is necessary to fix up the duration of the period say a week, a month, a quarter or year. This will depend upon the nature of the problem and the data available in each case. There are different methods of measuring seasonal variations but the following are more commonly used :—

(*a*) Monthly average method.
(*b*) Monthly averages corrected for trend.
(*c*) Link-relative method.
(*d*) Moving averages method.

(a) **Monthly Average Method** :— This is a simple method by which the averages for different months are ascertained from the data extending over a number of years.

Table—5. **Index Numbers of Variable Yield Securities in India** (Base 1927–28=100)

Months	Year					Seasonal variations		
	1939	1940	1941	1942	1943	Monthly average	Secular trend	Seasonal Index
(a)	(b)	(c)	(d)	(e)	(f)	(g)	(h)	(i)
January	105.4	120.9	121.2	128.5	152.7	125.7	124.4	101.0
February	103.6	121.6	122.9	121.7	158.3	125.6	125.6	100.0
March	100.8	122.1	123.8	119.6	162.8	124.8	126.8	98.4
April	97.6	123.0	118.2	119.1	174.5	126.5	128.0	98.8
May	100.8	115.6	119.9	122.8	170.7	126.9	129.2	97.5
June	100.4	116.0	123.2	124.3	174.8	127.7	130.4	97.9
July	100.0	111.7	129.2	126.8	168.8	127.3	131.6	96.7
August	98.5	109.5	129.0	127.4	170.9	127.1	132.8	95.7
September	111.4	110.6	136.2	136.3	176.3	134.1	134.0	100.1
October	111.9	114.2	138.8	140.0	183.1	137.6	135.2	101.8
November	124.7	119.8	151.6	149.2	188.9	146.8	136.4	107.6
December	130.7	119.3	132.1	143.7	186.8	142.5	137.6	103.6
Total average	107.2	117.0	128.8	130.0	172.4	130.0	131.0	100.0

In table no. 5 figures for five years are given in columns (b), (c), (d), (e), (f) and the monthly averages are given in column (g). This is a very simple operation and can be easily understood. But in this the defect is that these monthly averages include the effects of the trend besides the seasonal variations. To avoid this the monthly averages are corrected with the trend figures as shown in the following method.

(b) **Monthly Averages Corrected for Trend**:—In this method the effect of the secular trend is eliminated by doing further operations after the monthly averages are computed. The secular trend is calculated by any convenient method. In table No. 6 the secular trend has been computed by the Least Squares Method. For the year it comes to 14.34 and for each month it is 1. 2. As there is an even number of items i. e. 12 in table 5 the average of 131.0 cannot be written against June but it will be half-way between the two. The monthly average of 1. 2 has, therefore, been divided by 2 to find the variation for half the month and it is added to 131.0 to have a figure for July (131.0 + 0.6=131.6) and it has been deducted from 131.0 to get the figure for June (131.0—0.6=130.4). The secular trend column has been written adjusting the difference of 1.2 for each month. The seasonal index column in table 5 has been worked out by finding the percentage of monthly average to the secular trend.

Seasonal Index is also worked out by calculating the percentage of the figure of the monthly average to the average of monthly averages. But the seasonal index as calculated in table 5 col. (i) is preferrable.

Table No. 6. Secular Trend (By Least Squares Method.)

Years (a)	Annual average (b)	Deviation (c)	Deviation. squared (d)	Product of cols· (b) and (c) (e)
1939	107.2	-2	4	-214.4
1940	117.0	-1	1	-117.0
1941	128.8	0	0	0
1942	130.0	1	1	130.0
1943	172.4	2	4	344.8
Total averege	655.4 131.0		10	143.4

143.4÷10=14.34 Annual.
14.34÷12=1.2 Monthly.

(c) **Link Relative Method:**—This method is no doubt one of the best methods of computing seasonal indices but the process is complicated and requires a number of calculations to be made. First of all, "Link relatives" are computed which are ratios or percentages obtained by dividing the figure of any month by that of the preceding one. A typical item usually a median or mean is selected which is then related to a common base. The next step is to convert these link relatives with a fixed base. Allowance is then made for the discrepancy caused by the trend. The corrected relatives are then expressed as percentages of their arithmetic averages. As said above this method is not easy for manipulation.

(d) **Moving Averages Method:**—Seasonal variations can also be calculated by the Moving Averages method the general principles of which have already been explained. For the computation of seasonal variation by this method, moving averages should be calculated as shown in table 7 columns (4) and (5). The actual data should then be turned into percentages of the moving averages as shown in column (6). The data should then be re-arranged as shown in table 8 and the percentage average as shown in table 8 column (5). The percentages in column (5) should then be converted into percentages based on the average obtained (vide bottom of this column) and written in column (6) as Seasonal Indices.

Months (1)	Actual sales (2)	Sums in 12's (3)	Twelve months moving average (4)	Two months average from col. (3) (5)	Actual sales as percentages of moving averages (6)
1942					
January	26				
February	24				
March	21				
April	19				
May	21				
June	22				
	. .	283	23.6	...	...
July	21	...	...	23.7	88.6
	...	285	23.8	...	...
August	20	...	...	23.8	84.0
	...	286	23.8	...	...
September	23	...	...	23.9	96.2
	...	288	24.0	...	...
October	25	...	...	24.0	104.1
	...	289	24.1	...	...
November	30	...	...	24.1	124.4
	...	290	24.2	...	...
December	31	...	. .	24.2	127.2
	...	291	24.3	...	...
1943					
January	28	...	...	24.3	115.2
	...	293	24.4	...	...
February	25	...	...	24.4	102.4
	...	293	24.4	...	...
March	23	...	...	24.4	94.2
	...	294	24.5	...	...
April	20	...	..	25.5	81.6
	...	294	24.5	...	...
May	22	...	...	24.6	89.4
	...	296	24.7	...	...
June	23	...	...	24.7	93.1
	...	296	24.7	...	...

Months (1)	Actual Sales (2)	Sums in 12's (3)	Twelve months moving average	Two months average from col. (3) (4)	Actual sales as precentage of moving averages (5)
July	23	...	...	24.7	93.1
		296	24.7	...	...
August	20	..	...	24.7	80.9
	...	297	24.8	...	...
September	24	...	...	24.8	96.7
	..	297	24.8	...	...
October	25	...	...	24.8	100.8
	...	298	24.8	...	...
November	32	...	. .	24.8	129.0
	. .	298	24.8	...	...
December	31	. .	...	24.8	125.0
	...	298	24.8	...	...
1944					
January	28	...	...	24.8	112.9
	...	297	24.8	...	...
February	26	...	...	24.8	104.5
	...	297	24.8	...	...
March	23	...	...	24.8	92.7
	. .	299	24.9	...	...
April	21	...	...	25.0	84.0
	...	301	25.1	...	...
May	22	...	...	25.1	87.6
	...	301	25.1	...	...
June	23	...	...	25.2	91.2
	.	304	25.3	...	...
July	22				
	...				
August	20				
September	26				
October	27				
November	32				
December	34				

Table 8.—**Sales in 1942–44.**

Months (1)	1942 (2)	1943 (3)	1944 (4)	Average of cols. 2,3 and 4. (5)	Seasonal index (6)
January		115.2	112.9	114.0	114.1
Feb		102.4	104.5	103.4	103.5
March		94.2	92.7	93.4	93.5
April		81.6	84.0	82.8	82.9
May		89.4	87.6	88.5	88.6
June		93.1	91.2	92.1	92.2
July	88.6	93.1		90.8	90.9
Aug.	84.0	80.9		82.4	82.5
Sept.	96.2	96-7		96.4	96.5
Oct.	104.1	100.8		102.4	102.5
Nov.	124.2	129.0		126.6	126.7
Dec.	127.2	125.0		126.1	126.2
				1198.9 ÷ 12 =99.9	

Really speaking to avoid cyclical influence it is advisable to take a longer period but for the sake of convenience the shorter period of three years has been taken in this example.

Exercises

1. What is a trend ? Name the important types of movements influencing a time series and describe fully any one of them.

2. **What is a long-time trend ? Compare and contrast it with seasonal or periodic variation.**

3. Explain the principal methods available for ascertaining the secular trend and state with reasons the method you prefer.

4. Describe the method of ascertaining the moving average and state the difficulty experienced in arriving at the moving average for even number of years.

5. Write short notes on;—

   Progressive average, Link relatives, Centering of the moving average, Tertiary variation and Cyclical movement.

6. Name some types of periodic fluctuations that are not of a seasonal character.

7. What is meant by a trend ? How would you statistically eliminate the influence of seasonal and cyclical factors in the long period movement of any series ? (1936–Mar.)

8. Calculate five-year moving averages of students reading in the Alpha Technical Institute from the following table.

Year	No. of students	Year	No. of students
1	328	8	427
2	317	9	405
3	357	10	438
4	392	11	445
5	402	12	447
6	405	13	480
7	410	14	482

9 **Find nine year moving averages of the number of mills in existence from 1913-14 to 1933-34. Plot these results.**

1913–14	264		
1914–15	255	1924–25	305
1915–16	267	1925–26	303
1916–17	267	1926–27	306
1917–18	269	1927–28	297
1918–19	264	1928–29	292
1919–20	265	1929–30	304
1920–21	253	1930–31	310
1921–22	271	1931–32	317
1922–23	289	1932–33	331
1923–24	310	1933–34	344

10 Calculate five-year moving averages of area under tea in India.

Year	Area in 1000 acres	Year	Area in 1000 acres
1925	672	1930	802
1926	679	1931	807
1927	690	1932	809
1928	702	1933	816
1929	712	1934	821
		1935	827

Plot on a piece of graph paper the annual area under cultivation as well as its five yearly moving averages.

11 **The following table shows total deposits of all exchange banks in India *in crores* of rupees. Calculate five year and nine-year moving averages and show all of them graphically.**

Year	Deposits	Year	Deposits
1915	33.5	1925	70.5
1916	38.0	1926	71.5
1917	53.4	1927	68.9
1918	61.9	1928	71.1
1919	74.4	1929	66.7
1920	74.8	1930	68.1
1921	75.2	1931	67.5
1922	73.4	1932	73.1
1923	68.4	1933	70.8
1924	70.6	1934	71.4
		1935	76.2

12. Show graphically the following :—

Results of working of class I railways in India :

(in millions of £)

	Total capital outlay	Gross earning
1923–24	464	60
1924–25	473	74
1925–26	487	73
1926–27	505	72
1927–28	594	86
1928–29	599	86
1929–30	617	84
1930–31	627	77
1931–32	631	71
1932–33	638	70
1933–34	635	72

Plot also quinquennial moving averages of the above in broken lines.

13. The following table shows the yearly available supplies of all cereals per adult equivalent of population in India for a number of years.

Year	Available supplies lb.
1931–32	427
1932–33	416
1933–34	408
1934–35	408
1935–36	379
1936–37	405
1937–38	390
1938–39	360
1939–40	382
1940–41	352
1941–42	355
1942–43	358
1943–44	383
1944–45	361

Draw a graph to represent the above data. Calculate 3–year moving averages and plot them on the same chart. What are the deviations of the actual values from the moving average for the year 1941–42 and 1943–44 ?

(1947-March.)

14. Draw a graph of the following time series and study its trend.

Year	Value	Year	Value
1910	496	1920	1442
1911	615	1921	1617
1912	686	1922	1678
1913	835	1923	1791
1914	888	1924	1916
1915	1081	1925	1883
1916	1132	1926	2064
1917	1139	1927	2278
1918	1320	1928	2368
1919	1389	1929	2345

15. The following table gives the value of exports of merchandise from India during the year 1919–20 to 1923–24. Calculate the seasonal variations for each month during this period.

In crores of rupees

Months	1919–20	1920–21	1921–22	1922–23	1923–24
April	20	27	17	23	29
May	20	26	18	26	28
June	19	21	15	18	29
July	26	19	17	23	25
August	25	19	18	24	22
September	30	21	19	20	23
October	28	19	17	21	25
November	29	17	19	27	26
December	26	18	21	26	30
January	29	18	22	28	36
February	26	17	21	30	35
March	30	18	26	31	40

16. Plot the following figures on a graph paper and study the trend.

Wholesale price index number

	Calcutta	Bombay
	(June 1914=100)	(July 1914=100)
1925	159	163
1926	148	149
1927	148	147
1928	145	146
1929	141	145
1930	116	126
1931	96	109
1932	91	109
1933	87	98
1934	89	95
1935	91	99
1936	91	96
1937	102	106
1238	95	101
1939	108	109

CHAPTER X

DISPERSION

When a mass of quantitative data is collected for a statistical purpose it is tabulated with a view to reducing it to a form from which its characteristics as a whole may be readily determined. A single significant expression is then computed to show the net result of the series or group and the various methods for ascertaining the different forms of such abbreviated expressions, better known as averages, have been explained at length already. This form of a summary expression is used for several purposes, one of which is for comparing two or more groups or series but an average alone is not a satisfactory criterion for such a purpose as will be seen from the following illustration.

Salaries of a dozen clerks in two offices

Office A	Office B
Rs.	Rs.
75	75
35	70
45	60
65	80
50	65
130	75
80	90
60	80
55	60
140	85
90	80
75	80
Total Rs. 900	Rs. 900
Average Rs. 75	Rs. 75

We have seen that the average salary of clerks in both the offices A and B is Rs. 75 but it is clear that the composition of the group is radically different. The variation from the average in group B is very much less as compared to that in group A. The fluctuations in the group A are from 35 to 130 while in the other they are only from 60 to 90. In the statistical language the dispersion is less in the case of group B or in other words this group is more uniform than the other. Dispersion, in brief. is lack of uniformity in the sizes of the items of a group or series and it is useful for ascertaining to what extent the average is typical.

Absolute and Relative Dispersion:—This diversion or deviation which gives an idea of spreading or scattering is ascertained to indicate the lack of uniformity in magnitudes but in matters of comparison the absolute measure of dispersion is likely to be misleading as will be seen from the following example.

Marks of students in two subjects

	English out of 100	Optionals out of 300
A	40	120
B	50	150
C	80	240
D	30	90
E	90	270
Total.	290	870
Average.	58	174

From this it will be seen that the percentage of marks of students A, B, C, D and E in English and in optionals is the same but the relation in the case of absolute number of marks being one to three. When, however, we compare the absolute dispersions in both the cases we find them to be very different. In the first case the marks vary from 30 to 90 while in the other the variability extends from 90 to 270. In the absence of careful study this gives an erroneous impression. This is avoided by computing relative measures of dispersion, which is obtained by dividing the absolute measure of dispersion by the average selected as the basis for calculation. The absolute measure, therefore, is quite insignificant when different class intervals as well as different units are used as compared to the relative measures or the coefficient of variation as it is usually called. Thus, measures of variation in the weights of gold and coal mines, if calculated in ounces, may show tremendously high degree of absolute variability in the case of coal but if the relative measures of dispersion are computed this extraordinary diversion will automatically disappear as will be seen from the detailed study of the different measures of dispersion.

Measures of Dispersion :—The various measures of dispersion commonly used are :—

1. The Range.
2. The Quartile Deviation.
3. The Mean Deviation
4. The Standard Deviation.
5. The Lorenz Curve.

The Range :—The computation of the dispersion by the Range method is very easy as it is obtained by finding the difference in size between the extreme items but for comparing the dispersion, the measure thus arrived at is not satisfactory as will be seen in the following cases of weights and heights of one hundred persons,

Weights in lb. of 100 persons

Class interval	Mid-value	No. of persons f	Deviations d	fd	Cum. frequncy cf
90–100	95	5	−4	−20	5
100–110	105	12	−3	−36	17
110–120	115	8	−2	−16	25
120–130	125	14	−1	−14	39
130–140	135	9	0	0	48
140–150	145	16	1	16	64
150–160	155	18	2	36	82
160–170	165	9	3	27	91
170–180	175	8	4	32	99
180–190	185	1	5	5	100
Total		100		30	

$$\text{Arithmetic average} = 135 + \frac{30}{100} \times 10 = 138 \text{ lb.}$$

$$\text{Median} = 140 + \frac{50.5-48}{16} \times 10 = 141.5$$

$$\text{Mode} = 150 + \frac{9}{16+9} \times 10 = 153.6$$

Heights in inches of 100 persons

Class interval	Mid-value	No. of persons f	Deviations	fd.	Cum. frequency cf
61–62	61.5	2	–4	–8	2
62–63	62.5	5	–3	–15	7
63–64	63.5	12	–2	–24	19
64–65	64.5	14	–1	-14	33
65–66	65.5	14	0	0	47
66–67	66.5	17	1	17	64
67–68	67.5	21	2	42	85
68–69	68.5	8	3	24	93
69–70	69.5	7	4	28	100
Total		100		50	

$$\text{Arithmetic average} = 65{\cdot}5 + \frac{50}{100} \times 1 = 66 \text{ inches}$$

$$\text{Median} = 66 + \frac{50.5-47}{17} \times 10 = 68.06$$

$$\text{Mode} = 67 + \frac{8}{17+8} \times 10 = 70.2$$

The average weight is 138 lb. and it ranges from 90 to 190, the absolute range being 100 lb. The average height happens to be 66 and the range is from 61 to 70 giving only 9 as the absolute range. We are, therefore, likely to think that the dispersion in the case of weights is disproportionately high. This is erroneous as will be seen if we calculate relative measures of dispersion.

Relative measure :—This is obtained by the conversion of the absolute measure by comparing it with some standard and for this, a fractional expression is created by dividing the absolute dispersion by the standard used which is usually the arithmetic average of the group or by double the value of this average. It is not usual to find the median or the mode used in place of the arithmetic mean to obtain the relative dispersion.

Relative dispersions of range

	Weights	Heights
By using arithmetic average	$\frac{190-90}{138} = 0.72$	$\frac{70-61}{66} = 0.14$
By using double the arithmetic average	$\frac{190-90}{138 \times 2} = 0.36$	$\frac{70-61}{66 \times 2} = 0.07$
By using median	$\frac{190-90}{141.5} = 0.70$	$\frac{70-61}{68.06} = 0.13$
By using mode	$\frac{190-90}{153.6} = 0.65$	$\frac{70-61}{70.2} = 0.12$

In the case of absolute range the variation in the first case is 100 and in the second only 9 thus showing a very wide divergence but by using the relative measure of dispersion the difference is found to be much less comparatively.

Another method of obtaining the relative measure of dispersion in case of the range is to divide the absolute measure by the sum of the extremes which in the above case will be as under :—

Weights $\frac{190-90}{190+90} = \frac{100}{280} = 0.35$

Heights $\frac{70-61}{70+61} = \frac{9}{131} = 0.07$

Characteristics of the Range method :—The range as a measure of dispersion commands very little importance. The chief defect in this measure is that it is based on the observations of the two extremes and no weight is attached to other items even though their position in the group is of very great importance. Again the two extreme items taken for computation are such as are likely to be influenced by abnormal occurences. In spite of these defects this method is often used but mostly by those who are not familiar with statistics.

The Cumulative Range :—As the simple range as a measure of dispersion is not of much value the cumulative range is used to convey a summarized idea of the data. This is done by computing the cumulative frequency and then calculating the percentage as shown in the following table :—

Heights of 100 persons in inches

Height	Frequency	Cum. frequency	Percentage of the total frequency
61–62	2	2	2.0
62–63	5	7	7.0
63–64	12	19	19.0
64–65	14	33	33.0
65–66	14	47	47.0
66–67	17	64	64.0
67–68	21	85	85.0
68–69	8	93	93.0
69–70	7	100	100.0

The Quartile Deviation or Semi-interquartile Range :— While considering the utility of the range as a measure of dispersion it is noticed that it gives too much importance to the extreme items. This measure of the quartile deviation is an improvement on it as it is based on the upper and lower quartiles which are not likely to be as erratic as the extreme items. The median shows the magnitude of the item just halfway while the quartiles are the values of items one-fourth of the distance from the two extremes. As the interval between the two quartiles comprises of the central half of the items of the group this measure is an improvement upon the previous one. The formula for computing the absolute measure of dispersion by the Quartile Deviation method is as under :—

$$Q.\ D. = \frac{Q_3 - Q_1}{2}$$

The Quartile deviations in the case of the two examples of weights and heights are :-

Weights	Heights
$\frac{Q_3-Q_1}{2}$	$\frac{Q_3-Q_1}{2}$
$= \frac{156.53-120.18}{2}$	$= \frac{67{\cdot}56-64.45}{2}$
$= 18.17$	$= 1.55$

Its old name is ***Probable Error*** and the rough rule is that the semi-interquartile range is approximately two thirds of the standard-deviation.

Relative measure :—The relative measure of dispersion in this case is computed by dividing the absolute measure by median or by the average of two quartiles. The latter method is, however, more common.

Relative Dispersion of Quartile Deviation

Based on	Weights	Heights
(1) Median $\dfrac{\frac{Q_3-Q_1}{2}}{\text{Median}}$	$\dfrac{\frac{156.53-120.18}{2}}{141.5}$ $= 0.12$	$\dfrac{\frac{67.56-64.45}{2}}{68.06}$ $= 0.02$
(2) The average of quartiles $\dfrac{\frac{Q_3-Q_1}{2}}{\frac{Q_3+Q_1}{2}}$	$\dfrac{\frac{156.53-12.18}{2}}{\frac{156.53+120.18}{2}}$ $= 0.13$	$\dfrac{\frac{67.56-64.45}{2}}{\frac{67.56+64.45}{2}}$ $= 0.02$

Characteristics of the Quartile Deviation method :—This measure of dispersion is easy to calculate as well as to comprehend. It is no doubt based on a few more items as compared with the range but even then it fails to be sufficiently influenced by many items in the group. It is not a measure of variation from any particular form of average and is not capable of further mathematical treatment.

The Mean Deviation :—In ascertaining the range as well as the quartile deviation measures were computed by taking into consideration the two extremes in the case of the former and the two quartiles in the case of the latter. In no case the deviations from any particular point were taken for calculation. Mean Deviation, which is a better measure of dispersion is the arithmetic mean of the deviations of the series from some fixed point in the calculation of which ***algebraic signs are ignored.*** This fixed point for ascertaining the deviations can be mean, median, or mode. Theoretically median is given more importance as the value of the mean deviation is the least when it is based on the median while in practice the use of arithmetic mean is common.

When the mean deviation is to be calculated from a frequency distribution table it is assumed that all the items in a particular class are concentrated at the mid-point of the class and so the mid-value of the class is taken for computation. The mean Deviation is to be computed by the formula :—

$$\text{M. D.} = \frac{\Sigma f d}{n}$$

Where Σ = the sum (ignoring the signs)

f = the frequency

d = the deviations of the individual items from the average selected (mean, median or mode)

n = the total number of items.

Computations of Mean Deviation *based on Median.*

Marks of 100 students of section A.

Marks	Mid value	Frequency f	Cumulative frequency	Deviation from median 'd'	fd.
10–20	15	19	19	−30.4	−577.6
20–30	25	3	22	−20.4	−61.2
30–40	35	2	24	−10.4	−20.8
40–50	45	49	73	−0.4	−19.6
50–60	55	24	97	9.6	225.6
60–70	65	2	99	19.6	39.2
70–80	75	0	99	29.6	0
80–90	85	1	100	39.6	39.6
	Total	100			−679.2 +304.4

$$\text{Median} = 40 + \frac{50.5-24}{49} \times 10 = 45{\cdot}41$$

$$\text{Mean Deviation} = \frac{679.2+304.4}{100} = \frac{983.6}{100} = 9.836$$

Computation of Mean Deviation *based on Mode.*

Marks of 100 students of section A.

Marks	Mid-value	Frequency f	Deviation from mode 'd'	fd.
10–20	15	19	−34.2	−649.8
20–30	25	3	−24.2	− 72.6
30–40	35	2	−14.2	− 28.4
40–50	45	49	− 4.2	−195.8
50–60	55	24	5.8	139.2
60–70	65	2	15.8	31.6
70–80	75	0	25.8	0
80–90	85	1	35.8	35.8
Total		100		−946.6 +206.6

$$\text{Mode} = 40 + \frac{25}{2+24} \times 10 = 49.23$$

$$\text{Mean Deviation} = \frac{946.6 + 206.6}{100} = \frac{1153.2}{100} = 11.532$$

Computation of Mean Deviation Based on *Mean*

Marks of 100 students of section A.

Marks	Mid-value	Frequency f	Mid-point × f	Deviation from mean (d)	fd.
10–20	15	19	285	−26.7	−507.3
20–30	25	3	75	−16.7	− 50.1
30–40	35	2	70	− 6.7	− 13.4
40–50	45	49	2205	3.3	161.7
50–60	55	24	1320	13.3	319.7
60–70	65	2	130	23.3	46.6
70–80	75	0	0	33.3	0
80–90	85	1	85	43.3	43.3
Total		100	4170		−570.8 +570.8

$$\text{Arithmetic mean} = \frac{4170}{100} = 41.7$$

$$\text{Mean Deviation} = \frac{570.8 + 570.8}{100} = \frac{1141.6}{100} = 11.416$$

Relative Measure :—The mean deviation of 100 students of section A has been computed separately by three methods based on median, mode and mean. The relative measure of dispersion will in each case be computed by dividing the absolute measure by the respective average used for calculating the deviation in the cases concerned.

Relative Dispersion of Mean Deviation

$$\text{Based on Median} = \frac{\text{M. D.}}{\text{Median}} = \frac{9.8}{45.4} = 0.21$$

$$\text{Based on Mode} = \frac{\text{M. D.}}{\text{Mode}} = \frac{11.5}{49.2} = 0.23$$

$$\text{Based on mean} = \frac{\text{M. D.}}{\text{Mean}} = \frac{11.4}{41.7} = 0.27$$

Characteristics of the Mean Deviation Method:—The deviation as a measure of dispersion is, unlike the range or the quartile deviation, based on all items and is, therefore, affected by the values of all of them, It is not unduly affected by the extreme items but it is defective as in the computation of this measure algebraic signs are disregarded.

Short-cut Method:—From the working of the last example wherein the deviations from the actual mean have been taken it is clear that the method is labourious because of the treatment of fractions. This can be simplified if an arbitrary integer is taken for calculating the deviations and the necessary correction made later on as shown in the following formula :—

$$\text{M. D.} = \frac{\Sigma fd + c(n_1 - n_2) - q}{n} \times i$$

Where Σfd = total of the fd column disregarding the algebraic signs.

c = the total of the fd column divided by the total number of observations. (*Signs to be taken into account.*

n^1 = the number of observations below the mean

n^2 = the number of observations above the mean

i = the class interval.

n = the total number of observations

q = double the amounts of fd of items falling between the classes covered by the mean and arbitrary mean. (*Signs not to be taken into account.*)

Computation of Mean Deviation *based on mean.*

Marks of 100 students of section A

Marks	Mid-value	frequency	Deviations from the arbitrary mean	fd.
10–20	15	19	−3	−57
20–30	25	3	−2	−6
30–40	35	2	−1	−2
40–50	45	49	0	0
50–60	55	24	1	24
60–70	65	2	2	4
70–80	75	0	3	0
80–90	85	1	4	4
Total		100		−65 32

$$\text{Arithmetic mean} = 45 + \frac{(-65+32)}{100} \times 10 = 41.7$$

$$\text{Mean Deviation} = \frac{(65+32) + \frac{(-65+32)}{100} \times (24-76) - 2 \times 0}{100} \times 10$$

$$= \frac{97 + (-0.33)(-52)}{100} \times 10$$

$$= \frac{97 + 17.16}{10}$$

$$= 11.416$$

In this example the mean is 41.7 and so the number of items below the mean includes the frequencies of all the classes having a mid-value of less than 41.7 and so n^1 is 24 being the sum of 19, 3, and 2. In the same way n^2 includes the frequencies of classes having the mid–value of more than 41.7 and comes to 76. The other way is to deduct n_1 from n i. e. 100–24. The letter 'C' stands for $\frac{\Sigma fd}{n}$ i. e. $\frac{(-65+32)}{100}$ As there are no items falling between the classes covered by the mean and the arbitrary mean 'q' comes to 2×0 i. e. 0.

The mean deviation in this case has been based on the mean. If it is to be based on median or mode the word median or mode will be substituted in place of the mean in the procedure outlined above.

The Standard Deviation—In computing the mean deviation algebraic signs were disregarded but this difficulty is overcome in the case of standard deviation by taking the squares of the deviations as a negative quantity when squared becomes positive. This measure of deviation is more extensively used and is calculated by taking the deviations usually from the arithmetic mean though they can also be taken from the median or the mode.

The standard deviation which is also known as '*Error of Mean Square*', '*Mean Square Error*', '*Root–mean square deviation*' or '*Mean Error*' is the square root of the sum of the squares of the individual deviation from the arithmetic mean divided by the total number of observations. The name root–mean square deviation is very appropriate as it describes the method. This can also be computed by taking the deviations from the median or the mode but it is customary to have them from the mean as the sum of square of such deviations is the minimum. The standard deviation is shown by the Greek letter sigma and its formula is :—

$$\text{S.D. or } \sigma = \frac{\Sigma fd^2}{n}$$

where Σ = the sum of
f = the frequency
d = deviations of individual items from the mean.
n = the total number.

The quantity $(S.D.)^2$ is known as the *Variance*.

Computation of the standard deviation.

Heights of 140 Persons

Height in inches	Mid–value	Frequency f	Deviation from the mean d	fd.	fd2
59–60	59.5	1	–5	–5	25
60–61	60.5	5	–4	–20	80
61–62	61.5	12	–3	-36	108
62–63	62.5	21	–2	–42	84
63–64	63.5	23	-1	–23	23
64–65	64.5	20	0	0	0
65–66	65.5	24	1	24	24
66–67	66.5	14	2	28	56
67–68	67.5	11	3	33	99
68–69	68.5	4	4	16	64
69–70	69.5	5	5	25	125
Total		140		-126 +126	688

$$\text{Mean} = 64.5 + \frac{126-126}{140} \times 1 = 64.5$$

$$\text{S.D.} = \sqrt{\frac{688}{140}} = \sqrt{4.91} = 2.21$$

Relative Measure :—As the standard deviation has been computed by taking deviations from the mean the relative measure of dispersion or the coefficient is based on the mean. Relative dispersion of $\text{S.D} = \frac{\text{S. D.}}{\text{Mean}} = \frac{2.21}{64.5} = 0.03$

If the deviations are taken from the median or the mode for ascertaining the relative dispersion, the standard deviation is to be divided by the median or the mode as the case may be and not by the mean.

Characteristics of Standard Deviation :—Standard deviation as a measure of dispersion is of great importance. It is free from the defect of the mean deviation in the calculation of which the algebraic signs are disregarded. It is not only based on all the items in the group, but each one of them is given its due weight. It can be computed with reasonable ease and can be used in the further mathematical treatment. As it is rigidly defined, there is no scope for personal judgment which very often varies. No doubt comparatively more calculations are involved in its computation, but looking to the other points it is one of the best measures of dispersion and is, therefore, most commonly used.

Short-cut Method:—The computation of the standard deviation from the frequency table of the heights of 140 students did not involve many calculations, particularly of the fractions as the mean is an integer, but such cases are rare, To get over this difficulty a mid-point of a class near the centre of the distribution is selected as an arbitrary mean and the standard deviation is then calculated with the formula slightly modified to suit this change. The deviations in this case are taken from the arbitrary mean and the correction is made by the insertion of '$-c^2$'.

$$S.D = \sqrt{\frac{\Sigma fd^2}{n} - c^2} \times 1$$

where $c = \frac{\Sigma fd}{n}$ i. e. the total of the fd column divided by the total number of observations.

The result of the previous example can be verified by the use of this formula as under :—

Computation of Standard Deviation.

Heights in inches of 140 persons

Height	Mid-value.	Frequency.	Deviations from arbitrary mean.	fd	fd^2
59–60	59.5	1	–4	–4	16
60–61	60.5	5	–3	–15	45
61-62	61.5	12	–2	–24	48
62–63	62.5	21	–1	–21	21
63–64	63.5	23	0	0	0
64–65	64.5	20	1	20	20
65–66	65.5	24	2	48	96
66–67	66.5	14	3	42	126
67–68	67.5	11	4	44	176
68–69	68.5	4	5	20	100
69–70	69.5	5	6	30	180
Total		140		–64 +204 140	828

$$S.D. = \sqrt{\frac{828}{140} - \left(\frac{140}{140}\right)^2} \times 2$$

$$\times 1$$

$$= \sqrt{5.91 - 1}$$

$$\times 1$$

$$= \sqrt{4.91}$$

Relations between different measures of dispersion:— The range is the distance between the two extremes of the scale and so all the observations are covered by it. The range equal to twice the quartile deviation includes one-half of the total number of items as it is the distance on each side of the midway point between the two quartiles. In a fairly symmetrical distribution, a range of 7½ times the mean deviation or six times the standard deviation includes approximately 99% of the total observations. The empirical rule based on the above is that in fairly distributed series the mean deviation is very nearly four-fifths of the standard deviation. This can be verified by calculating the standard deviation of the frequency giving the marks of 100 students of section A the mean deviation of which as already computed by taking deviations from the mean is 11.4.

Computation of Standard Deviation.

Marks of 100 students of section A

Marks	Mid-value	Frequency f	Deviations from the arbitrary mean d.	fd	fd²
10–20	15	19	–3	–57	171
20–30	25	3	–2	–6	12
30–40	35	2	–1	–2	2
40–50	45	40	–0	0	0
50–60	55	24	1	24	24
60–70	65	2	2	4	8
70–80	75	0	3	0	0
80–90	85	1	4	4	16
Total		100		–65 32	233

$$\text{Mean} = 45 + \frac{-65+32}{100} \times 10 = 41.7$$

$$\text{S. D} = \sqrt{\frac{233}{100} - \left(\frac{-65+32}{100}\right)^2} \times 10$$

$$= \sqrt{2.33 - 0.11} \times 10 \qquad = \sqrt{2.22} \times 10$$

$$= 1.49 \times 10 = 14.9$$

The mean deviation is 11.4 while the standard deviation is 14.9 and so $\frac{\text{mean deviation}}{\text{standard deviation}} = 0.76$. The series in question is not fairly symmetrical and hence this slight variation in the result. The range obtained by multiplying the mean deviation by 7½ is 11.4 × 7½ i. e. 85.5 and it includes all the observations. The same is the case if we take the range by multiplying the standard deviation by six. $(14.9 \times 6 = 89.4)$

It is very necessary to give due attention to the measures of dispression as it is possible that though the average of both the frequency distribution tables is the same, there may be a great divergence in the dispersion. This is illustrated by the following two tables of the marks of the students of two different sections I and II.

Marks of students of section II.

Marks	Mid-value	Frequency f	Deviations d	fd	fd^2
10–20	15	37	-3	-111	333
20–30	25	17	-2	-34	68
30–40	35	3	-1	- 3	3
40–50	45	8	0	0	0
50–60	55	1	1	1	1
60–70	65	1	2	2	4
70–80	75	20	3	60	180
80–90	85	13	4	52	208
Total		100		-148 115	797

$$\text{Arithmetic mean} = 45 + \frac{-148+115}{100} \times 10 = 41.7$$

$$\text{Mean deviation} = \frac{(148+115) - \frac{(57-43)(-33)}{100}}{100} \times 10$$

$$= \frac{263 - 4.62}{10} = \frac{268.38}{10} = 25.838$$

Coefficient of M. D. $= \frac{25.838}{41.7} = 0.619$

Standard deviation $= \sqrt{\frac{797}{100} - \left(\frac{-33}{100}\right)} \times 10$

$$= \sqrt{7.97 - 0.11} \times 10$$

$$= \sqrt{7.86} \times 10$$

$$= 28.0$$

Coefficient of S. D. $= \frac{28.0}{41.7}$

$$= 0.67$$

Marks of students of section I.

Marks	Mid-value	Frequency f	Deviation d	fd	fd^2
10–20	15	12	–3	–36	108
20–30	25	20	–2	–40	80
30–40	35	15	–1	–15	15
40–50	45	30	0	0	0
50–60	55	7	1	7	7
60–70	65	5	2	10	20
70–80	75	3	3	9	27
80–90	85	8	4	32	128
Total		100		–91 58	385

Arithmetic mean $= 45 + \frac{(58-91)}{100} \times 10 = 45 - 3.3$

$= 4.17$

Mean deviation $= \frac{(91+58) + \frac{(47-53)(-33)}{100}}{100} \times 10$

$= \frac{149+1.98}{10} = \frac{150.98}{10} = 15.098$

Coefficient of M. D. $= \frac{15.098}{41.7} = 0.36$

Standard Deviation $= \sqrt{\frac{385}{100} - \left(\frac{-33}{100}\right)^2} \times 10$

$= \sqrt{3.85 - 0.11} \times 10$

$= \sqrt{3.74} \times 10$

$= 19.0$

Coefficient of S. D. $= \frac{19.0}{41.7}$

$= 0.45$

Coefficient of Variation :—One of the other measures of dispersion or scatter used with some modification is the coefficient of variation and it is based on standard deviation and the arithmetic mean. The formula as suggested by Professor Karl Parson for a relative measure of dispersion is denoted by

$$\text{'V' is } \frac{100 \times \text{Standard Deviation}}{\text{Arithmetic Mean.}}$$

By using this formula for the marks of section I we get:—

$$V = \frac{100 \times 19.0}{41.7} = 45.5$$

and for section II

$$V = \frac{100 \times 28.0}{41.7} = 61.1$$

By comparing the coefficient of variation of the two sections I and II it is evident that the variation is more in the marks of section II and the same has been shown by the comparison of the figures of the coefficients of Mean and Standard Deviation.

It will be seen that the average of marks in both the sections I and II is 41.7 but their dispersions are different as will be clear from the following table :—

	Section I	Section II
Arithmetic mean	41.7	41.7
Mean deviation	15.1	25.8
Standard deviation	19.0	28.0
Coefficient of M. D.	0.36	0.62
Coefficient of S. D.	0.45	0.67

The Lorenz Curve:—The Lorenz curve is a cumulative frequency graph designed to show the deviations in the distribution of the particular subject of enquiry over the group of the data. As the method was originally devised by Dr. Max. O. Lorenz for measuring the evenness of the distribution of wealth, it is known as Lorenz Curve and is extensively used in the study of the distribution of wealth, income, wages, etc.

The Lorenz Curve is drawn by plotting the respective points of the percentages of the cumulative frequencies on a graph paper upon which 100 per cent deviation of the axis of X i. e. the horizontal scale coincides with the zero percent division of the vertical scale i. e. the axis of Y. The scale on the axis X is not marked in the usual manner *but from the right to left*. The two extreme points are joined by a straight line which represents the case having absolutely no dispersion. The Lorenz Curve shows the extent of dispersion—the more the area between this straight line and the curve the more is the extent of dispersion.

Data for the Lorenz Curve

Income for ten years.

Years	Rs.	Years	Rs.
1931	3300	1936	7200
1932	4700	1937	2900
1933	4900	1938	7800
1934	2400	1939	4000
1935	5000	1940	7800

From this information, a separate table is to be prepared giving the equal percentage for each year as shown in column (2), the cumulative frequency of the data and its percentage for each year worked on the total as in column (4). The Lorenz Curve is then drawn based on the points plotted from columns (2) and (4) The straight line joining the two ends, represents the case having absolutely no dispersion with which this Lorenz Curve drawn will be compared to see the extent of dispersion.

Table–1 **Data for the Lorenz Curve**

Year (1)	Percent (2)	Income Cumulative frequency (3)	Income Percentage of the total. (4)
1931	10	3300	6.6
1932 and prior	20	8000	16.0
1933 and ,,	30	12900	25.8
1934 and ,,	40	15300	30.6
1935 and ,,	50	20300	40.6
1936 and ,,	60	27500	55.0
1937 and ,,	70	30400	60.8
1938 and ,,	80	38200	76.4
1939 and ,,	90	42200	84.4
1940 and ,,	100	50000	100.0

The Lorenz Curve

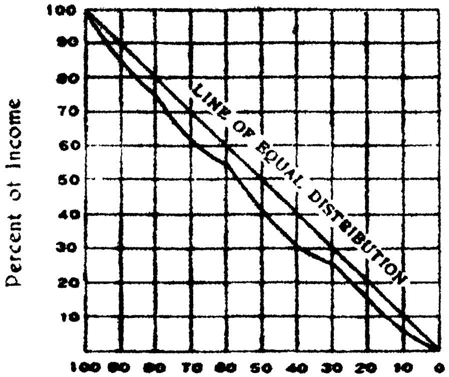


Fig. 1. Percent years.

In the above case, only one Lorenz Curve is drawn and so there is no scope for comparison between two curves which is possible in the example given below.

Table–2 **Salaries of clerks in firms A and B**

Salaries	Firm A	Firm B
Not exceeding Rs. 40	25	48
Exceeding Rs. 40 but not exceeding Rs. 80	30	32
„ 80 „ „ 120	25	25
„ 120 „ „ 160	50	43
„ 160 „ „ 200	26	22
„ 200 „ „ 240	12	20
„ 240 „ „ 280	22	5
„ 280 „ „ 320	10	5
	200	200

Table-3 Data for the Lorenz Curve

Salaries			Firm A		Firm B	
Amount 1	Average 2	Percent 3	Cum. frequency 4	Percent 5	Cum. frequency 6	Percent 7
Less than Rs. 40	20	6.6	25	12.5	48	24.0
„ „ 80	60	20.0	55	27.5	80	40.0
„ „ 120	100	33.3	80	40.0	105	52.5
„ „ 160	140	46.6	130	65.0	148	74.0
„ „ 200	180	60.0	156	78.0	170	85.0
„ „ 240	220	73.3	168	84.0	190	95.0
„ „ 280	260	86.6	190	95.0	195	97.5
„ „ 320	300	100.0	200	100.0	200	100.0

The Lorenz Curves

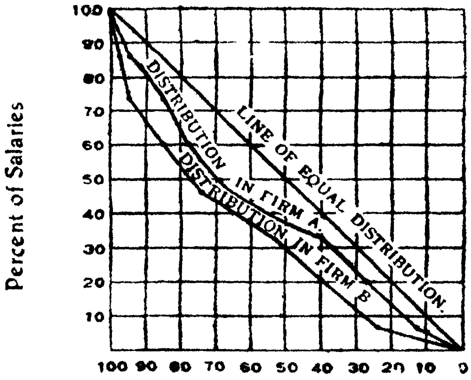


Fig 1 Percent of firm

Points will be plotted by taking figures from columns 3 and 5 and then by taking them from 3 and 7. Two curves will thus be drawn to facilitate comparsion of the dispersion.

Characteristics of the Lorenz Curve;—The Lorenz Curve is very useful in showing the extent of dispersion graphically. It, however, fails to give any numerical measurement of the distribution for comparative purposes. The method of writing the scale on the axis of X is unusual and the curves almost concide at the extreme ends making it difficult to study them minutely. To avoid this, the extremes are again plotted on a separate graph sheet taking a much larger scale.

The Moments of Dispersion:—Usually dispersion is ascertained by taking the average deviation from mean, mode or median. The moments are used for this purpose as will be seen from the following description of the moments which are three is number.

(1) **The First Moment**:—It is also known as the average or mean deviation and is the sum of the deviations from the average divided by the number of observations i. e. $\frac{\Sigma fd.}{n}$

(2) **The Second Moment**:—This moment is the sum of the deviations when squared and divided by the number of observtions i. e. $\frac{\Sigma fd^2}{n}$. The square root of this is the standard devition.

(3) **The Third Moment**:—As the second moment is obtained by squaring the deviations, this third moment is the sum of the cubes of the deviations i. e. $\frac{\Sigma fd^3}{n}$

The first, second and third moments are often shown by the symbols v_1, v_2, and v_3 respectively and the most significant moments used in statistics are those based on the arithmetic mean.

The Modulus:—It is a measure of dispersion rarely used in elementary statistics and is obtained by using the following formula:—

$$\text{The Modulus or } c = \frac{\sqrt{2\Sigma f d^2}}{n}$$

The Modulus worked out from the example of the heights in inches of 140 students given under the head of standard deviation is

$$\text{Modulus} = \sqrt{2 \times \frac{688}{140}} = \sqrt{9.82} = 3.1$$

Exercise

1. What is dispersion ? Give the important measures for computing it and state with reasons which method you prefer.
2. Discuss the value of quartile deviation as a measure of dispersion.
3. Describe the method of computing standard deviation and state the usual ratio between this and mean deviation.
4. Explain by giving suitable examples the need of prefering relative measures of dispersion to absolute measures.
5. Explain clearly what you understand by dispersion and coefficient of variation. (1935–March).
6. Write short notes on:—Second moment, modulus, cumulative range and variance.
7. Describe the method of drawing Lorenz Curve and state its utility in statistical work.
8. From the following table giving heights of students, calculate Quartile Deviation and also Standard Deviation:—

Height in inches	No. of students	Height in inches	No. of students
53	15	61	8
55	11	63	14
57	18	65	12
59	10	67	9
		69	3

2. Find mean and standard deviations of examination marks of 75 students.

Marks	No. of students getting them	Marks	No. of students getting them.
20	2	55	7
25	4	60	5
30	7	65	9
35	5	70	8
40	3	75	4
45	8	80	3
50	6	85	2
		90	2

10. Compare the following distributions by the aid of the median and the quartiles and state your conclusions in words.

Sizes of Spinning Firms.

Spindles (figures in 000s).	Number of firms. 1884	1921	Spindles (figures in 000s).	Number of firms. 1884	1921
0–10	27	6	90–100	4	18
10–20	47	15	100–110	3	26
20–30	25	6	110–120	3	13
30–40	15	8	120–130	1	7
40–50	17	14	130–140	1	8
50–60	18	11	140–150	1	6
60–70	16	19	150–160	...	3
70–80	21	16	160–170	...	1
80–90	2	22	170–180	...	1
			180–190	...	2

(1943–Oct.)

11. **Calculate mean and standard deviations of the following data and give the ratio of mean deviation to standard deviation :—**

Wages per month Rs.	50	60	70	80	90	100	110	120	130	140	150
No. of recipients.	22	27	15	18	23	14	31	22	18	7	3

12. The following table gives the imports and exports of India : calculate a measure of dispersion and state in which the variation is greater.

	Crores of rupees.			Crores of rupees.	
Year.	Exports.	Imports.	Year.	Exports.	Imports.
1927–28	319	250	1932–33	239	203
1928–29	339	263	1933–34	275	182
1929–30	345	258	1934–35	280	210
1930–31	308	206	1935–36	282	216
1931–32	263	176	1936–37	342	199

Adopt another measure of dispersion and verify the result.

13. Compare the variations of height and weight of boys from the following :—

9339 boys of age 5 years had a mean height of 41.26 inches, standard deviation 2.238 inches, their mean weight was 39.724 lb, standard deviation 4.335 lb. (1935–Mar).

14. Compare the variation in scores of two batsmen A and B in first class cricket matches in 1936 from the following data of ten innings :—

A 53, 27, 61, 39, 7, 105, 80, 16, 5, 56.

B 21, 16, 78, 83, 41, 43, 57, 95, 4, 0. (1937–Mar.)

15. Compare the inequality of holdings in the two Districts 'A' and 'B' as disclosed by the following data :—

Size of holding	Number of cultivaters in District 'A'	District 'B'
Upto 5 acres	61,514	1,67,311
Over 5 and below 15 acres	80,042	73,273
,, 15 ,, ,, 25 ,,	44,474	19,858
,, 25 ,, ,, 100 ,,	19,072	9,024
,, 100 ,, ,, 500 ,,	1,673	244
,, 500 acres.	31	1

(1943–Mar.)

16. What purposes does a measure of dispersion serve ? Define standard deviation and semi-interquartile range and calculate them from the following table giving the age distribution of 542 members of the House of Commons :—

Age	Number of members.	Age	Number of members.
20–30	3	60–70	140
30–40	61	70–80	51
40–50	132	80–90	2
50–60	153		

(1943–Mar.)

17. **Find the median, mean deviation and standard deviation for the following distribution of tomato plants :—**

No. of Tomatoes per plant.	No. of plants.
0	2
1	5
2	7
3	11
4	18
5	24
6	12
7	8
8	6
9	4
10	3

(1946–Oct.)

18. Find the mean yield of paddy and the standard deviation for the distribution of the results of 3061 crop cutting experiments shown in the following table :—

Yield of paddy per acre in lb.	No. of experiments.
0–400	236
401–800	481
801–1200	604
1201–1600	576
1601–2000	419
2001–2400	333
2401–2800	217
2801–3200	87
3201–3600	64
3601–4000	23
4001–4400	14
4401–4800	6
4801–5200	1
Total	3061

(1945–Oct.)

19. Frequency distributions of cost of production of gur (in annas per maund) for different holdings in two districts are given below :—

Cost	District A No. of holdings.	District B No. of holdings.
32–48	9	1
48–64	32	10
64–80	37	34
80–96	21	23
96–112	13	21
112–128	7	14
128–144	5	10
144–160	2	9
160–176	1	5
176–192	2	2
192–208	1	1
Total	130	130

In which district is the average cost greater and by how much ? In which district are costs more variable ?

(1946–Mar.)

20. The fluctuations in rates of Kohinoor and Tata Deferred on the 7th March are given below. Find out which of the two shares shows greater variability.

Kohinoor—618, 619, 616, 623, 620, 624, 622, 625, 622, 625, 626, 625.

Tata Deferred—2152½, 2132½, 2134¼, 2132½, 2145, 2142½, 2146¼, 2130, 2146¼, 2130, 2146¼, 2142½, 2150, 2135, 2132½.

21. Compare the level and dispersion of wheat prices in the different divisions of the Bombay presidency from the following data :—

Wholesale prices of wheat per maund.

Northern division	Price Rs.	Central division	Price Rs.	Southern division	Price Rs.
Bombay	7–0–0	Ahmednagar	7–14–0	Belgaum	8–3–0
Ahmedabad	7–10–0	Jalgaon	6–14–6	Bijapur	7–0–0
Kaira	9–0–0	Dhulia	7–2–8	Dharwar	7–12–0
Broach	9–0–0	Nasik	6–12–0	Kanara	8–14–0
Surat	7–5–0	Satara	7–6–0	Kolaba	8–6–9
		Poona	8–4–6	Ratnagiri	8–13–0
		Sholapur	7–15–8		

(1920 Oct.)

22. Find Mean and Standard Deviations of the distribution given below :—

No. of accidents.	Persons having said number of accidents.
0	15
1	16
2	21
3	10
4	17
5	8
6	4
7	2
8	1
9	2
10	2
11	—
12	2
Total	100

(1944–Oct.)

23. The following table gives the age distributions of students admitted to a college in the years 1914 and 1938. Find which of the two groups is more variable in age.

Age	Number of students admitted in	
	1914	1938
15–	—	1
16–	1	6
17–	3	34
18–	8	22
19–	12	35
20–	14	20
21–	14	7
22–	5	19
23–	2	3
24–	3	—
25–	1	—
26–	–	1
27–	1	—
Total	64	148

(1942–Mar.)

CHAPTER XI

SKEWNESS

From the preceding chapter it is clear that a single significant value in the form of an average representing the central tendency of the distribution is not quite enough to give a complete description of the series. It is necessary to have other values as well to measure the chief characteristics of the distribution and to compare it with others. We have already realised the utility of the measure of dispersion from the central value and the second measure of the degree of symmetry will be obtained by the study of skewness which will indicate the tendency of the group to depart from the symmetry or the form of the distribution.

Skewness is a lack of symmetry and it indicates in the case of a frequency distribution curve that it is turned more to one side than to the other and as such it has a longer tail on one side. If the observations in the group have been symmetrically placed i. e. at points of equal deviations above and below the mode and if the frequencies are equal the curve constructed will have a bell-shaped form as shown by the curve B in fig. 1. Such an ideal curve is rarely met with and it will be skewed either to the right as shown by the curve C, or to the lett as in the case of the curve A. The skewness to the right i. e. having a lower range to the right is shown by a positive sign (+), and that to the left by a negative sign (−). The positive sign occurs when the skewness exists between the median and the third quartile in which case the right side is elongated and the curve takes the form as shown by the curve C. When the elongation is towards the left, as in the case of the curve A the sign will be nagative, the skewness then occuring in the lower half of the array as will be seen from the curves drawn based on the table giving the earnings of the three groups of labourers.

Table-1. **Earnings of the groups of workers.**

Earnings.	Mid-values	Group A.	Group B.	Group C.
Below 35	30	1		
35—45	40	2		
45—55	50	3		
55—65	60	4		
65—75	70	6	2	
75—85	80	8	4	
85—95	90	14	7	
95-105	100	18	10	2
105-115	110	23	19	16
115-125	120	26	26	26
125-135	130	4	19	20
135-145	140	1	10	14
145-155	150		7	10
155-165	160		4	7
165-175	170		2	5
175-185	180			4
185-195	190			3
195-205	200			2
205-215	210			1
Total		110	110	110

The mode being the point of the greatest density is the top-point in the curve which can spread out in one of the three ways given above. The median bisects the area of the histogram and if the distribution is symmetrical the median will move towards the mode and in an absolutely symmetrical curve the median and mode will coincide. In the same way the arithmetic mean which is always located at the centre of gravity of the histogram would also coincide with the mode in an ideally symmetrical distribution as can be seen in the case of the curve B.

Dispersion and Skewness :—The dispersion is computed to measure the deviations from the type and in case of different measures the importance given differs. In the standard deviation more importance is given to items removed from the type chosen.

Earnings of group A, B and C.

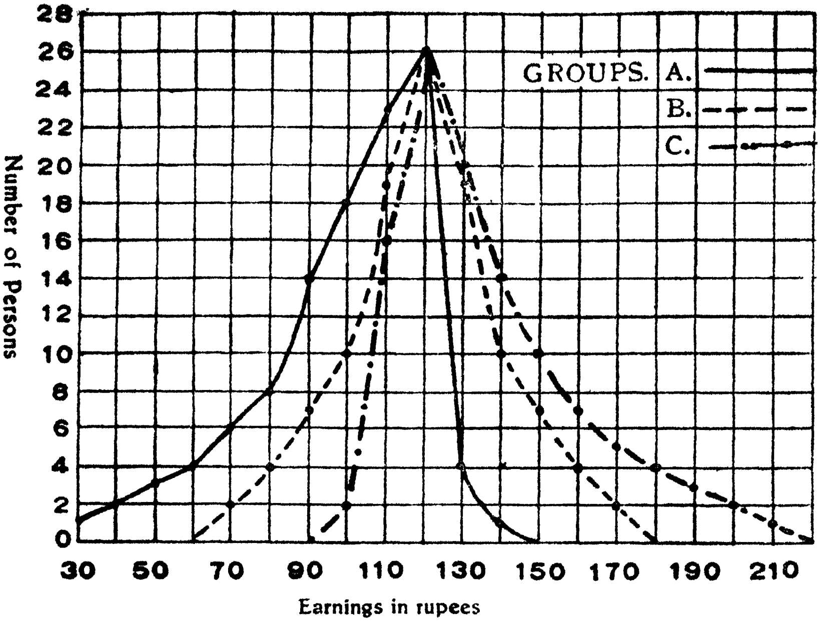


Fig. 1

while in the mean deviation no such emphasis is given. In the case of quartile deviation the differences included are only those that lie within the boundaries of the upper and the lower quartiles. The measures of dispersion in short fail to show the manner of the distribution of the deviation. Skewness, on the contrary, shows the position in relation to the mode or the median and brings out the distortion from the symmetrical form of the curve. The measures of skewness not only measure the relative or absolute skewness but also show the side on which the symmetry lies.

Measures of Skewness:—Skewness can be easily noticed by the graphic method i. e. by drawing a frequency distribution curve but it fails to give a numerical measure for comparing the skewness. This can be achieved by adopting various methods some of which are too complicated and hence the use of one of the following measures is recommended:—

1. The Position of averages method.
2. The Quartile deviation method.
3. The Cubed deviation method.

Relative Measure of Skewness:— The absolute measure of skewness will not be a proper measure for comparison for reasons already explained in the case of measures of dispersion and hence in each case a relative measure or the coefficient of skewness will have to be computed.

The Position of Averages Method:—In a perfact symmetrical distribution the mean, median and the mode coincide but the distance of the arithmetic mean and the median from the mode increases with the non-symmetric distribution. The absolute measure of skewness easily obtained is (mean-mode) and the relative measure or the coefficient of skewness for the purpose of comparison which is indicated by the symbol 'j' is computed by the following formula of Karl Pearson.

$$j = \frac{\text{Mean-Mode}}{\text{Standard Deviation.}}$$

The value of arithmetic mean will be greater than that of the mode in the case of a curve elongated to the right i. e. of the type C and hence the result will be positive.

If happens In some cases that the mode is not very clearly defined and then it becomes difficult to determine it. In such a case the following equation which is based on the approximate relation between mean, median and mode is made use of for substituting the value of the mode.

$$\text{Mode} = \text{Mean} - 3\,(\text{Mean} - \text{Median})$$

The formula, when the value of mode is substituted is

$$j = \frac{3\,(\text{Mean—Median})}{\text{Standard Deviation.}}$$

The other way of computing the skewness based on the same principle is by taking (Mean-Median) as the numerator but this will reduce the value of skewness to one-third as compared with the previous formula and when the skewness is slight this will fail to indicate more clearly the degree of skewness.

In place of standard deviation as a denominator mean deviation is also used and then the formula is

$$j = \frac{\text{Mean-Median}}{\text{Mean Deviation.}}$$

In this case mean deviation is based on arithmetic mean. It is to be based on median and not on mean if the following formula is to be used to avoid the indefiniteness of the mode.

$$j = \frac{\text{Mean-Mode}}{\text{Mean-deviation.}}$$

It is, however, necessary that the same measure of dispersion should be used as the denominator in all the cases when a comparison of skewness is involved in the case of a number of groups.

In the Position of the averages method Pearson's Coefficient of skewness occupies an important place and according to G. Undy Yule there is 'only one generally recognised measure of skewness and that is Pearson's measure.'

Characteristics of the Position of Averages Method:— Arithmetic mean, which is an essential item in the computation of skewness is no doubt based on all the items but mode is sometimes not properly defined. In such a case the median is substituted but it is only about one third the difference between mean and mode and hence the degree of contrast is lowered. This method is simple and fairly satisfactory in normal cases and hence this Pearson's measure is more commonly made use of.

The Quartile Deviation Method :—It is based on the principle that if the curve is asymmetrical the range from the first quartile to the median would not be the same as that from the third quartile to the median, or $(Q_3 - Me)$ will not be equal to $(Me - Q_1)$. The formula for the absolute measure of skewness by this method is skewness $(SK) = (Q_3 - Me) - (Me - Q_1)$ or briefly $Q_3 + Q_1 - 2Me$.

The relative measure or the coefficient in this method is obtained by dividing the absolute measure of skewness by quartile deviation.

Coefficient of skewness or $j = \dfrac{Q_3 + Q_1 - 2Me}{\dfrac{Q_3 - Q_1}{2}}$

Characteristics of the Quartile Deviation Method :— This measure of skewness like the quartile measure of dispersion fails to take into account the magnitude of extreme variations and really measures the skewness of the central one-half of the items.

It is, however, simple and easy to calculate and gives a fairly accurate result. Its lowest value is zero in the case of an ideally symmetrical dispersion. The maximum in a most asymmetrical distribution can go up to 2 though normally the numercial value dose not exceed 1. According to Dr. Bowley 0.1 indicates a moderate degree of skewness while 0.3 shows marked skewness. The value obtained by this method is not comparable with that ascertained by the Pearson's formula.

The Cubed Deviation Method :— This measure is ascertained from the cube root of the third moment of dispersion and is based on the principle that the sum of the deviations from arithmetic mean when the signs are considered is zero. If the items in the given group are symmetrical at points of equal deviation the frequencies are equal and so the result would be zero. If the deviations from arithmetic mean are squared all the signs would become positive and the distinction between positive and negative deviations will not exist. Cubes of such deviations will not only preserve this distinction of the positive and negative signs but the items at the extremes would get a relatively higher importance which will be of great value when the skewness is to be measured. The absolute measure of skewness will then be $\sqrt{\frac{\Sigma f d^3}{n}}$ and the relative measure or the coefficient of skewness by this method will be obtained by dividing the absolute measure by mean or standard deviation of the series, the use of the standard deviation is, however, more common.

$$j = \frac{\sqrt{\frac{\Sigma f d^3}{n}}}{S\,D.} \quad \text{or} \quad \frac{\sqrt{\frac{\Sigma f d^3}{n}}}{M.\,D.}$$

Characteristics of the Cubed Deviation Method:—This method is in one respect just the opposite of the previous method i. e. the quartile measure of skewness in which the size of the extreme variations is not taken into account while in this too much emphasis is given to the extreme items because of their being cubed. This method entails considerable amount of calculations and hence though its coefficient is the best the Pearson's method is more commonly used.

Table—2. **Earnings of workers in group A.**

Earnings	No. of workers f	Deviation d	fd	fd²	Cum. frequency
Below Rs. 35	1	−6	− 6	36	1
35–45	2	−5	−10	50	3
45–55	3	−4	−12	48	6
55–65	4	−3	−12	36	10
65–75	6	−2	−12	24	16
75–85	8	−1	− 8	8	24
85–95	14	0	0	0	38
95–105	18	1	18	18	56
105–115	23	2	46	92	79
115–125	26	3	78	234	105
125–135	4	4	16	64	109
135–145	1	5	5	25	110
Total	100		−60 +163	635	

$$\text{Mean} = 90 + \frac{163-60}{110} \times 10 = 99.36$$

$$\text{Median} = 95 + \frac{55.5-38}{18} \times 10 = 104.72$$

$$\text{First Quartile} = 85 + \frac{27.75-24}{14} \times 10 = 87.67$$

$$\text{Third Quartile} = 115 + \frac{83.25-79}{26} \times 10 = 116.63$$

$$\text{Mode} = 115 + \frac{4}{4+23} \times 10 = 116.48$$

$$\text{Standard Deviation} = \sqrt{\frac{635}{100} - \left(\frac{103}{110}\right)^2} \times 10 = 22.1$$

$$\text{'j' (Pearson's)} = \frac{99.36-116.48}{22.1} = -0.77$$

'j' (Quartile Deviation Method) =

$$\frac{116.63+87.67-2\times 104.72}{\frac{116.63-87.67}{2}} = -0.35$$

Table—3. Earnings of workers in group B.

Earnings	No. of workers f	Deviation	fd	fd^2	Cum. frequency
Rs. 65– 75	2	–5	–10	50	9
75– 85	4	–4	–16	64	6
85– 95	7	–3	–21	63	13
95–105	10	–2	–20	40	23
105–115	19	–1	–19	19	42
115–125	26	0	0	0	68
125–135	19	1	19	19	87
135–145	10	2	20	40	97
145–155	7	3	21	63	104
155–165	4	4	16	64	108
165–175	2	5	10	50	110
Total	110		+86 –86	472	

$$\text{Mean} = 120 + \frac{86-86}{110} \times 10 = 120$$

$$\text{Median} = 115 + \frac{55.5-42}{26} \times 10 = 120.19$$

$$\text{First Quartile} = 105 + \frac{27.55-23}{19} \times 10 = 107.39$$

$$\text{Third Quartile} = 125 + \frac{83.25-65}{19} \times 10 = 133.02$$

$$\text{Mode} = 115 + \frac{19}{19+19} \times 10 = 120$$

$$\text{Standard Deviation} = \sqrt{\frac{472}{110}} \times 10 = 20.7$$

$$\text{`j' (Pearson)} = \frac{120-120}{20.7} = 0$$

$$\text{By (Q. D. Method)} = \frac{133.02+107.39-2\times 120.19}{\frac{133.02-107.39}{2}} = 0.002$$

Table—4. Earnings of workers in group C.

Earnings	No. of workers f	Deviation d	fd	fd^2	Cum frequency
Rs. 95–105	2	−4	− 8	32	2
105–115	16	−3	−48	144	18
115–125	26	−2	−52	104	44
125–135	20	−1	−20	20	64
135–145	14	0	0	0	78
145–155	10	1	10	10	88
155–165	7	2	14	28	95
165–175	5	3	15	45	100
175–185	4	4	16	64	104
185–195	3	5	15	75	107
195–205	2	6	12	72	109
205–215	1	7	7	49	110
Total	110		−128 89 −39	643	

$$\text{Mean} = 140 + \frac{89-128}{110} \times 10 = 136.46$$

$$\text{Median} = 125 + \frac{55.5-44}{20} \times 10 = 130.75$$

$$\text{First Quartile} = 115 + \frac{27.75-18}{26} \times 10 = 118.75$$

$$\text{Third Quartile} = 145 + \frac{83.25-78}{10} \times 10 = 150.25$$

$$\text{Mode} = 115 + \frac{20}{20+16} \times 10 = 120.55$$

$$\text{Standard Deviation} = \sqrt{\frac{643}{110} - \left(\frac{-39}{110}\right)^2} \times 10 = 23.9$$

$$\text{'j' (Pearson's)} = \frac{136.46-120.55}{23.9} = 0.66$$

$$\text{By (Q. D. Method)} = \frac{150.25+118.75-2\times130.75}{\frac{150.25-118.75}{2}} = 0.47$$

By computing the coefficient of skewness of the earnings of the three groups A, B and C it is seen that there is no skewness in the case of the group B as the distribution is symmetrical, the coefficient of skewness being only 0.002 which is negligible. In the case of the group A the skewness is on the left side and hence the coefficient is –0.77. As the coefficient in the case of the group C is 0.47 and it is positive the skewness is on the right side and is comparatively less than that in the case of group A. This can be verified by studying the curves representing the earnings of these three groups.

Exercises

1. Define skewness and give the different methods of ascertaining it.
2. Explain the method of computing skewness by the quartile deviation method. Do you think this to be a satisfactory method ?
3. What is skewness and how does it differ from dispersion ?
4. Give Pearson's measure of skewness.
5. Find the coefficient of skewness from sizes of items given below—

Weight in lb.	No. of persons	Weight in lb.	No. of persons
85	15	105	15
90	18	110	7
95	25	115	28
100	19	120	12
		125	11

6. Find the coefficient of skewness from the following data :—

Heights of school boys at age 5.

Height in inches	No.	Height in inches	No.	Height in inches	No.
28	1	36	166	44	567
29	0	37	344	45	233
30	1	38	740	46	89
31	3	39	1167	47	27
32	8	40	1670	48	19
33	13	41	1614	49	4
34	40	42	1541	50	4
35	59	43	1028	51	1

(1935–Oct.)

7. Find out coefficient of dispersion and a coefficient of skewness from the following table giving wages of 230 persons and explain their significance.

Wages	No. of persons	Wages	No. of persons
70–80	12	110–120	50
80–90	18	120–130	45
90–100	35	130–140	20
100–110	42	140–150	8

8. From the following table compute the quartile deviation as well as the coefficient of skewness.

Size	Frequency	Size	Frequency
2– 4	5	12–14	12
4– 6	12	14–16	10
6– 8	17	16–18	8
8–10	27	18–20	3
10–12	15		

9. How do you recognize the positive or negative character of skewness ? From the following table state in which section the coefficient of skewness is higher.

Marks Obtained.	Students of sec. A.	Students of sec. B.
10–20	5	6
20–30	12	9
30–40	20	14
40–50	30	42
50–60	40	30
60–70	7	9
70–80	5	7
80–90	1	3
Total	120	120

10. Weekly earnings of two groups of workers in factory A and B are given and you are asked to represent the matter graphically to see skewness and also to compute skewness by Pearson's method.

Weekly wages Rs.	No. of workers of factory A.	No. of workers of factory B.
8–12	5	30
22–16	6	35
15–20	8	40
20–23	10	50
24–28	25	70
22–32	30	35
32–36	46	15
36–40	50	13
40–44	60	12
44–49	70	10
Total	310	310

CHAPTER XII

CORRELATION

A mass of statistical matter is tabulated and single significant expressions like mean, median and mode are computed to give a summary of the data and to use them for making comparisons. Based on these, dispersions of items from the type are calculated to study the uniformity or otherwise of the group. As these fail to indicate the symmetry or divergence from the norm a coefficient of skewness is ascertained. Usually in the case of economic, social and scientific phenomena this study is not sufficient and a comparison of groups and series is required to be made with a view to seeing if there exists correlation, meaning thereby that between the two series or groups there exists some connection. For example, if there is a fairly well distributed rainfall in the country the prices of agricultural produce go down or students who are good at mathematics are usually weak in languages. It is our usual experience that if the supply of an article is reduced without any change in its demand the price goes up. All these go to show that a certain movement or tendency in the one produces an effect on the other which may be direct or inverse. In short, whenever two variables are so related that a movement or change in one is followed directly or inversely by a movement or change in the other and the greater the magnitude of the change in the one the greater is the magnitude of the change in the other the variables are said to be correlated. This relation is also called *Association* or *Co-variation.* The elements of the theory of correlation are contained in the investigations of the French astronomer Bravais but its development in the present form is mainly due to Galton's studies on Heredity in 1877.

Subject and Relative :—So far we have dealt primarily with the charactertistics of a single variable such as the central tendency and the variability as well as the symmetry of distribution. Correlation statistics are concerned with ascertaining the relationship berween variables and it is embodied in the theory of correlation. Out of the two series taken the one that is to be used as a standard or measure is called the ***Subject*** while the other which is to be compared or measured is known as the ***Relative.*** The variable that is dependent on the other is the dependent variable while the other is called 'Independent.'

Positive and Negative Correlation:— A variable, which is anything that can have a series of values such as weight in pounds, length in inches, may indicate relation which may be direct or indirect. If the supply of a commodity remains the same while the demand increases the price will rise and the correlation in this case is direct or positive. On the other hand if the demand remains unchanged while the supply increases the price will go down and this is then a case of negative or inverse correlation. If the frequency distribution curves are drawn in the case of a positive correlation the curves will be moving in the same direction but in the case of a negative correlation they will fluctuate in opposite directions. The numerical measure will be negative in the case of inverse and positive in the case of a direct correlation.

Degree of Correlation:—It is not quite sufficient to simply trace correlation between the variables but it is essential to ascertain the degree of its existence. The area of a square bears a definite relation to its side and so does the weight of water to its volume. In these cases correlation is positive as well as perfect. But correlation in the case of the supply and the price of a commodity is not likely to be perfect but it will be partial and negative. There will again be cases like the rainfall in inches and the number of births or the production of wheat and the manufacture of cloth in which there is no possibility of correlation. When we consider this question we may have cases ranging from perfect positive correlation to perfect negative correlation and in between will be cases of varying degrees of partial correlation, positive as well as negative, and also of its total absence. Partial correlation may again be high, moderate or low depending upon the degree of its relationship.

The degree of correlation can be measured by calculating a coefficient of correlation. It is, however, necessary to note that there may be by chance a numerical or pictorial indication of correlation in some cases but it will have to be seen if the movements in one set of data are really influenced by the movements in the other or that both sets of data are affected by another factor which is related to both. Wrong conclusions are likely to be drawn in some cases and for this purpose it is the duty of a statistician to guard against the influence of '*nonsense correlation*' as it is sometimes called.

Lag and Lead:— In the case of a time series it often happens that fairly long a period elapses between the cause and its effect with the result that though the two cycles agree they are not quite coincident in point of time. A close resemblance in fluctuations can in such cases be seen if an allowance is made for a lapse of time. For example, if rainfall is satisfactory prices of agricultural produce will not go down to a considerable extent immediately but it will take a few months to produce this result. When, therefore, the cycles do not quite coincide in point of time the one which follows is said to show a 'lag' of a given interval or a 'lead' by one that precedes. In the above case of rainfall and its effect on prices the time series for prices is said to lag behind that of rainfall. This very phenomenon can be seen in the case of other series too. Money spent on advertising and propaganda may not immediately show its effect on the sale of goods advertised. Of course, in matters of time series this occurrence is more common.

The lag and the lead can be easily determined by drawing curves for both the series and then by carefully marking the interval by which one curve can be shifted horizontally with a view to bringing the peaks and troughs of the two series on the same point of time. The easy method for doing this is to draw the curve of one series on a transparent graph paper and then to move the paper horizontally by placing it on the graph of the other series. The other way is to study the tabulated data carefully to see if an increase or decrease in one is followed by an increase or decrease in the other after allowing a lapse of time. While doing this it is necessary to bear in mind the possibility of negative correlation in which case the change visible will be in inverse direction. The standard mathematical method for ascertaining the amount of lag is to calculate coefficients of correlation between the two time series for different amounts of lag. It is necessary to note that a graphic representation does not correctly reveal which of the two curves is the cause and which the effect and so a careful study of the phenomena is necessary. Again, it will not be correct to assume the existence of a lag simply by looking at the curves and without carefully studying the cause and the effect of the particular phenomena. It should be noted that the term 'Lag' is also used to show that there exists a smaller degree of variation in one series than in the other.

Methods of Ascertaining Correlation:—In studying correlation it is necessary to bear in mind the distinction between the correlation of two 'series' and that of two 'groups' as all the methods may not equally suit both the cases. In case of series emphasis is laid on the time element and a comparison is made of fluctuations of two series for a certain period of time such as a comparison of the production of cotton and its prices say for a dozen years or of the imports and exports for some years. To the classification based on the element of space or on the possession or destitution of certain attributes or characteristics the name 'group' is given and a comparison of two groups of facts is made at a given time such as heights and weights of different persons or wages of persons working in cotton and jute mills.

The important methods adopted for ascertaining correlation are as under:—

1. The Graphic Method.
2. The Scatter or Dot Diagram Method.
3. Pearson's Method.
4. The Concurrent Deviation Method.
5. Spearman's Method of Ranks.

The Graphic Method :—This is purely a visual method of tracing the evidence of correlation. We can surely ascertain if the relation indicated by the flucuations is throughout the whole course, whether the maxima and minima occur at same points and whether the rise and fall is direct or inverse; but beyond establishing a relationsnip we cannot with the help of this method offer a definite measure for comparison. As compared with the other methods this is far less elaborate and its working can be easily understood. In this, curves are plotted from the data and then effort is made to comprehend visually if there exists any casual relation direct or inverse between curves as they stand, due allowance being made for the possibility of a lag.

If there exists positive correlation the curves will run more or less parallel (vide Fig. 2) but if it is negative they will be moving in opposite directions (vide Fig. 3). Such ideal conditions are rarely met with but there are cases of partial correlation and its degree (which may be high, moderate or low) is to be judged from their movements. If there is absence of correlation the curves will indicate no relation even after the possibility of lag is considered. As stated already it is necessary to study the various factors relating to the phenomena before deciding the existence of correlation.

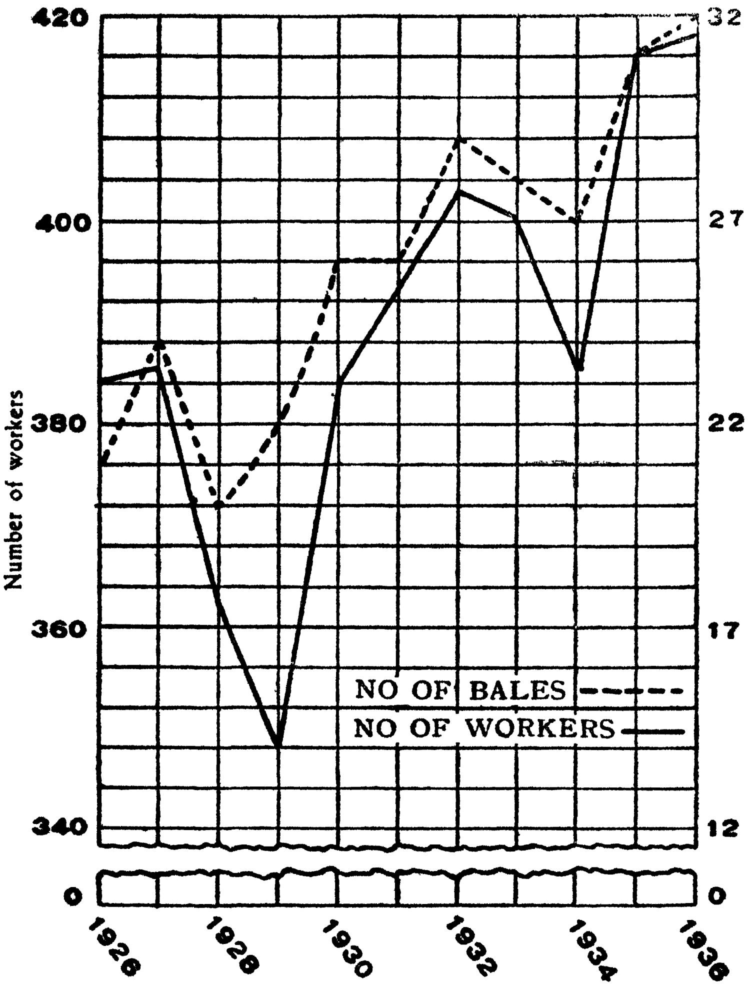
The graphic method is applicable to both series and groups without making important changes in the procedure as will be illustrated by the two examples (i) one of time series comparing the average number of workers employed daily and of bales consumed for a period of eleven years and (ii) the other of groups relating to the marks of twelve students in English and Mathematics.

For ascertaining the evidence of correlation between the above two time series the data will have to be represented by curves, one showing the number of workers employed and the other indicating the number of bales consumed. The number of workers in this case ranges from 348 to 418 and that of bales from 20 to 32. If both curves are plotted on the same scale it will have to be very high and in that case the fluctuations in the case of the second curve will not be properly visible as almost all the observations will come almost in a straight line. To get over this difficulty two different scales suitable to the series will have to be selected besides using the device of a 'false line' as it will allow us to adopt a more suitable scale for plotting the two curves. In selecting the different scales it is necessary to see that the scales are properly proportioned and are not such as will minimise the fluctuations in one series and exaggerate those in the other. For this purpose it is advantageous to equate the scales on the basis of their respective arithmetic average. It will be seen in the following graph that the lines of averages in both the cases are near each other. If they are absolutely close, these curves may overlap and thus spoil the graphic display.

Table—1 **Number of Workers and of Bales Consumed.**

Years	Average number of workers employed daily (in thousands)	Number of bales consumed (in lakhs)
1926	384	21
1927	385	24
1928	362	20
1929	348	22
1930	384	26
1931	395	26
1932	403	29
1933	400	28
1934	385	27
1935	415	31
1936	418	32

Number of Workers and of Bales Consumed.



By carefully studying the fluctuations of curves it is seen that though they are on the same side they are not absolutely parallel but a close relation can be visually comprehended. When we consider the number of workers and of the consumption of bales, a relation of positive nature is easily established. From the movements of the two curves which are fairly in sympathy it can be easily said that there exists partial correlation, the degree indicated being not low but high. No numerical measure of degree can, as already stated, be obtained by this method.

Plotting the Indices:— In using the above method there is the danger of using different scales which is likely to influence the result either adversely or favourably. To guard against this, absolute quantities can be turned into index numbers based on the arithmetic average of the series and then plotting the curves based on the indices thus obtained. In this case two different scales will not be required and the graphic display is sure to be more representative of the facts. The semi-logarithmic diagram will, however, be still more desirable and it can be prepared by plotting the logarithms of the absolute numbers or by using the logarithmic scale graph paper. Ordinarily the indices method is quite suitable.

Table—2 **Number of Workers and Bales Consumed**

Year	Average no. of workers employed daily		Bales consumed	
	Thousands	Indices	In lakhs	Indices
1926	384	98.7	21	80.8
1927	385	98.9	24	92.3
1928	362	93.1	20	76.9
1929	348	89.5	22	84.6
1930	384	98.7	26	100.0
1931	395	101.6	26	100.0
I932	403	103.6	29	111.6
1933	400	102.8	28	107.7
1934	385	98.9	27	103.8
1935	415	106.7	31	119.2
1936	418	107.5	32	123.1
Total average	4279 389		286 26	

In the first series the average is 389 and taking it as 100 indices have been calculated for the number of workers for each year. The arithmetic average of bales is 26 and by the same method indices have been worked out. The curves plotted from the indices also reveal a high degree of positive correlation as has been visible from the previous graph based on absolute quantities.

Indices of Number of Workers and Bales.

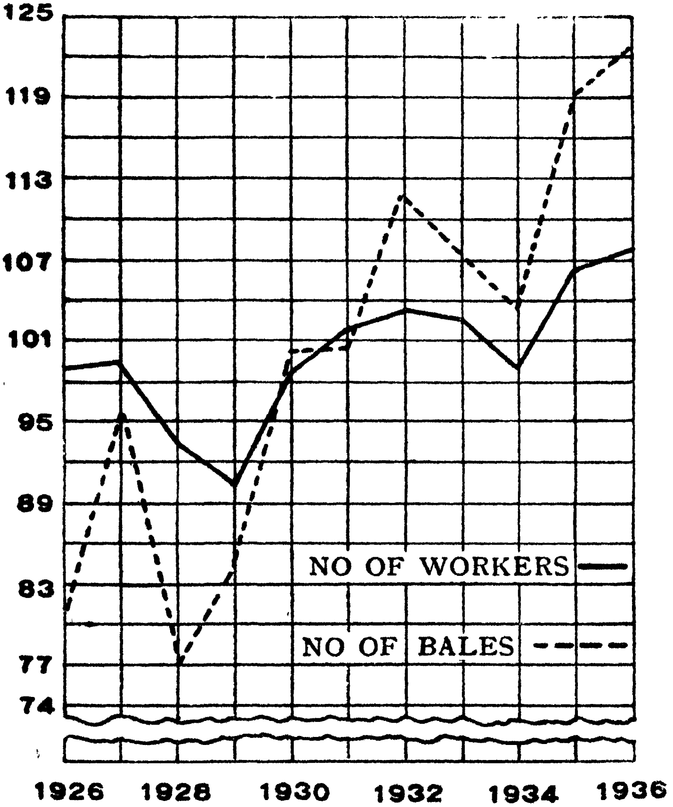


Fig. 2

Table—3 **Marks in English and Mathematics.**

Students	English	Mathematics
A	75	30
B	40	55
C	30	68
D	85	18
E	33	70
F	90	15
G	25	80
H	10	85
I	95	10
J	65	35
K	55	45
L	33	65

For plotting the curves for this example it is not necessary to have different scales or to use the device of a false line. On the horizontal line are marked points for students as A, B, C, D and on the vertical line are shown the marks. The movements of the two curves one representing marks in English and the other in mathematics, will be carefully studied to ascertain the existence of correlation. It can be seen that there is a definite relation of a very high degree but it is not direct.

Marks obtained in English and Mathematics

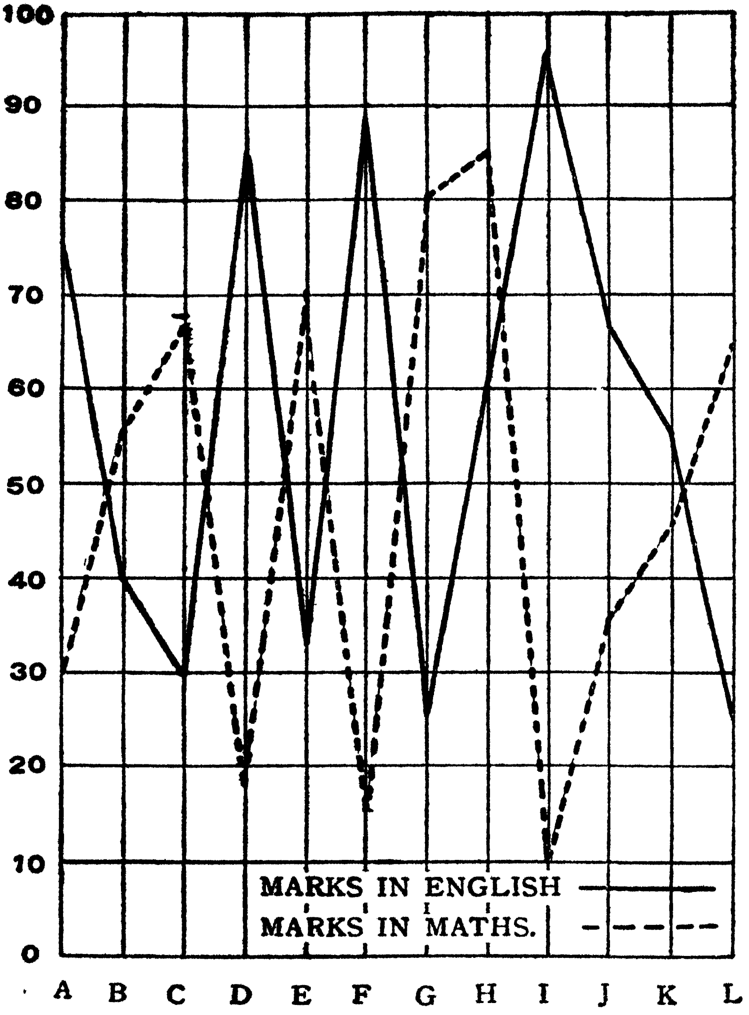


Fig. 3

Scatter or Dot Diagram:—The evidence of correlation can be ascertained by a Scatter or Dot diagram which is a chart prepared to exhibit graphically the relation between the two variables.

It is a very easy but rough method of finding out correlation by graphic device. The points are plotted on a piece of graph paper by taking convenient scales for the two frequencies and in case of the existence of correlation the plotted points will tend to concentrate in a band of greater or smaller width according to its degree. The greater the degree of correlation the narrower will be the band. In the absence of correlation the dots in the diagram will appear scattered. In the case of a perfect positive correlation the plotted points will form a straight line running between the lower left corner and the upper right. If it is negative the line will be running between the left hand upper corner and the lower right.

By plotting the data in Table 3 the following figure of a scatter diagram will be formed and as it is a case of more or less perfect negative correlation the band of points is narrow and is running from the left to right.

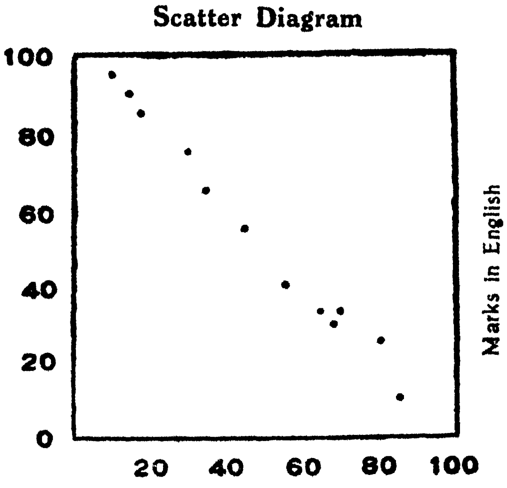
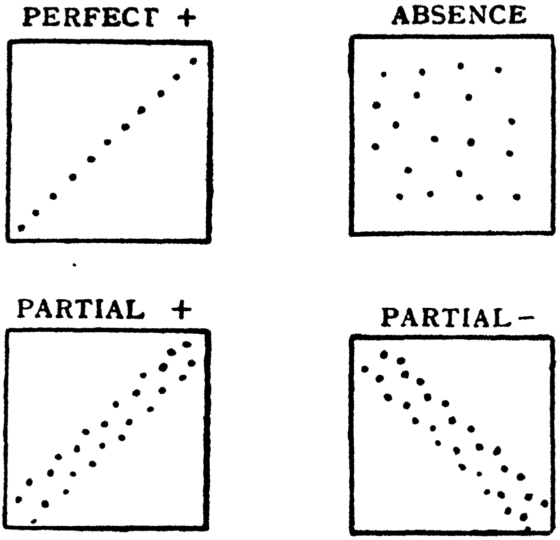


Fig. 4
Marks in mathematics.

The following diagram will illustrate the position of points in the other types of correlation.



The Pearson Method:—No numerical measure is obtained for making comparisons if the graphic method is made use of. This difficulty is got over by the mathematical formula devised by Karl Pearson, known as Pearson's Coefficient of Correlation and is indicated by the letter 'r'. As it gives a fairly accurate measure of correlation existing between variables it is most commonly used.

$$r = \frac{\Sigma xy}{N \times Sx \times Sy}$$

where x = deviations of subject items from their mean.
y = deviations of relative items from their mean.
N = Number of pairs of items.

S_x = Standard deviation of the subject.

S_y = Standard deviation of relative.

For the application of this formula the subject series should be written first and then the relative. Deviations from the arithmetic average for each item in both the series should be calculated and written in respective columns marked 'X' and 'Y'. Their squares should then be computed and in the last column should be entered the multiplication of items in 'X' and 'Y' columns. After standard deviations for both the series are obtained the coefficient can be computed by the formula given already. The probable error should also be calculated as it will be of great use in interpreting the result.

Probable Error:—In the case of a coefficient of correlation its reliability will depend upon the number of pairs compared for computing it and also its size. If more pairs are taken for comparison the result will be more reliable. The following formula is used for dealing with the error due to a limited number of items taken and the small size of the coefficient.

$$\underset{\text{(p. e.)}}{\text{Probable Error}} = \frac{0.6745\,(1 - r^2)}{\sqrt{N}}$$

Where r = the coefficient of correlation, and.

N = the number of pairs of items.

The Interpretation of the Coefficient of Correlation:—The degree of reliability of the coefficient is judged with the help of the probable error. If 'r' is less than the probable error the possibility of correlation is absolutely out of question but if the former exceeds six times the size of the latter there exists almost a certainty of correlation. If 'r' is less than 0.25 there is less evidence while if it is above 0.5 there is a decided evidence of correlation. Perfect correlation exists when the result is unity and whether it is positive or negative is to be judged from the algebraic sign of the coefficient.

Table–4 **Number of Workers and of Bales Consumed.**

Years	No. of hands in thousands	Deviations from mean x	x^2	Bales consumed (in lakhs)	Deviations from mean y	y^2	xy
1926	384	–5	25	21	–5	25	25
1927	385	–4	16	24	–2	4	8
1928	362	–27	729	20	–6	36	162
1929	348	–41	1681	22	–4	16	164
1930	384	–5	25	26	0	0	0
1931	395	6	36	26	0	0	0
1932	403	14	196	29	3	9	42
1933	400	11	121	28	2	4	22
1934	385	–4	16	27	1	1	–4
1935	415	26	676	31	5	25	130
1936	418	29	841	32	6	36	174
Total Mean	4279 389		4362	286 26		156	723

$$Sx = \sqrt{\frac{4362}{11}} = \sqrt{396.5} = 19.9$$

$$Sy = \sqrt{\frac{156}{11}} = \sqrt{14.2} = 3.8$$

$$r = \frac{723}{11 \times 19.9 \times 3.8} = \frac{723}{831.8} = 0.87$$

$$\text{Probable error} = \frac{0.6745-(0.87^2)}{\sqrt{11}}$$

$$= \frac{0.16}{3.3} = 0.05$$

$$r = 0.87 \pm 0.05$$

As 'r' is more than six times the size of the probable error and also above 0.5 there is a decided evidence of high degree of positive correlation.

The coefficient of correlation is also computed by using the following formula :—

$$r = \frac{\Sigma xy}{\sqrt{\Sigma x^2 \times \Sigma y^2}}$$

By using this formula for the above example we get

$$r = \frac{723}{\sqrt{4362 \times 156}}$$

$$= \frac{723}{824.9} = 0.87$$

The coefficient of correlation in the case of the above time series of the average number of workers employed daily and of bales consumed has been computed without any modification which is so necessary when long-time and short-time changes are to be given proper importance. The difference in the procedure to be adopted is mainly in ascertaining the deviations as will be seen from the example of index number of prices of raw cotton and cloth worked out by taking both the cases.

Long-time Changes :—When a correlation between long time trends of the two series is to be considered moving average figures will be ascertained for both the series and they will be taken as the original data. The deviations will then be calculated in each case from the arithmetic average of the moving averages and not from the original figures.

Table 5—Coefficient of Correlation—Long-time Changes.

Date	Raw cotton			
	Index number	Moving average	Deviation from average x	x^2
1935				
April 5	95	–	–	–
12	100	–	–	–
19	99	99.6	–0.9	0.81
26	102	101.2	0.7	0.49
May 3	102	102.0	1.5	2.25
10	103	103.2	2.7	7.29
17	104	103.6	3.1	9.61
24	105	103.6	3.1	9.61
31	104	103.2	2.7	7.29
June 7	102	102.8	2.3	5.29
14	101	102.4	1.9	3.61
21	102	102.4	1.9	3.61
28	103	102.8	2.3	5.29
July 5	104	103.6	3.1	9.61
12	104	103.6	3.1	9.61
19	105	103.0	2.5	6.25
26	102	101.6	1.1	1.21
Aug 2	100	100.4	–0.1	0.01
9	95	98.4	–2.1	4.41
16	98	96.6	–3.9	15.21
23	95	95.0	–5.5	30.25
30	93	94.2	–6.3	39.69
Sept 6	92	94.1	–6.3	39.69
13	93	94.4	–6.1	37.21
20	98	–	–	–
27	96	–	–	–
Total		2211.8		
Average		100.5		248.3

$$Sr = \sqrt{\frac{248.3}{22}} = \sqrt{11.28} = 3.4$$

$$Sy = \sqrt{\frac{10\ 84}{22}} = \sqrt{0.49} = 0.7$$

$$r = \frac{44.66}{22 \times 3.4 \times 0.7} = \frac{44.66}{52.36} = 0.85$$

Index Number of Prices of Raw Cotton and Cloth (July 1914)

Cloth.				
Index Number	Moving average	Deviation from average y	y^2	xy
110	...	...	—	–
111	...	...	—	–
111	111.2	–0.2	0.04	0·18
112	111.6	0.2	0.04	0.14
112	111.8	0.4	0.16	0.60
112	112.0	0.6	0.36	1.62
112	112.0	0.6	0.36	1.86
112	111.8	0.4	0.16	1.24
112	111.6	0.2	0.04	0.54
111	111.4	0	0	0
111	111.2	–0.2	0.04	–0.38
111	111.2	–0.2	0.04	–0.38
111	111.4	0	0	0
112	111.8	0.4	0.16	1.24
112	112.0	0.6	0.36	1.86
113	112.4	1.0	1.00	2.50
113	112.2	0.8	0.64	0.88
112	112.0	0.6	0.36	–0.06
111	111.6	0.2	0.04	–0.42
111	111.0	–0.4	0.16	1.56
111	110.4	–1.0	1.00	5.50
110	110.0	–1.4	1.96	8.82
109	110.0	–1.4	1.96	8.82
109	110.0	–1.4	1.96	8.54
111	...	—	–	–
111	...	—	–	–
	2450.6 111.4		10.84	44.66

$$P.E. = \frac{0.6745\,(1-0.85^2)}{\sqrt{22}} = \frac{0.19}{4.69} = 0.04$$

$$r = 0.85 \pm 0.04$$

Short Time Changes;—In this case the moving average figures will be calculated as in the case of long-time changes but the

Table-6. **Coefficient of Correlation**–*Short time changes.* **Index**

Date	Raw Cotton			
	Index number	Moving average	Deviations from moving average x	x^2
1935				
April 5	95	...	...	...
12	100	...	...	...
19	99	99.6	–0.6	0.36
26	102	101.2	0.8	0.64
May 3	102	102.0	0	0
10	103	103.2	–0.2	0.04
17	104	103.6	0.4	0.16
24	105	103.6	1.4	1.96
31	104	103.2	0.8	0.64
June 7	102	102.8	–0.8	0.64
14	101	102.4	–1.4	1.96
21	102	102.4	–0.4	0.16
28	103	102.8	0.2	0.04
July 5	104	103.6	0.4	0.16
12	104	103.6	0.4	0.16
19	105	103.0	2.0	4.00
26	102	101.6	0.4	0.16
August 2	100	100.4	–0.4	0.16
9	97	98.4	–1.4	1.96
16	98	96.6	1.4	1.96
23	95	95.0	0	0
30	93	94.2	–1.2	1.44
Sept. 6	92	94.2	–2.2	4.84
13	93	94.4	–1.4	1.96
20	98	...	...	...
27	96	...	...	...
Total				23.20

deviations in each case will be ascertained by taking the difference between the original data and the moving average figures. When long-time changes are considered these deviations are calculated by taking the difference between the arithmetic average of moving average figures and moving averages and not from the original data.

Number of Prices of Raw Cotton & Cloth (July 1914=100)

Cloth				xy
Index number	Moving average	Deviation from moving averages y	y^2	
110	...	...	...	...
111	...	...	...	...
111	111.2	–0.2	0.04	0.12
112	111.6	0.4	0.16	0.32
112	111.8	0.2	0.04	0
112	112.0	0	0	0
112	112.0	0	0	0
112	111.8	0.2	0.04	0.28
112	111.6	0.4	0.16	0.32
111	111.4	–0.4	0.16	0.32
111	111.2	–0.2	0.04	0.28
111	111.2	–0.2	0.04	0.08
111	111.4	–0.4	0.16	–0.08
112	111.8	0.2	0.04	0.08
112	112.0	0	0	0
113	112.4	0.6	0.36	1.20
113	112.2	0.8	0.64	0.32
112	112.0	0	0	0
111	111.6	–0.6	0.36	0.84
111	111.0	0	0	0
111	111.4	0.6	0.36	0
110	110.0	0	0	0
109	110.0	–1.0	1. 0	2.20
109	110.0	–1.0	1. 0	1.40
111	...	...	...	...
111	...	...	...	...

$$Sx = \sqrt{\frac{23.20}{22}} = \sqrt{1.05} = 1.03$$

$$Sy = \sqrt{\frac{4.60}{22}} = \sqrt{0.21} = 0.46$$

$$r = \frac{7.68}{22 \times 1.03 \times 0.46} = \frac{7.68}{10.42} = 0.73$$

$$\text{P. E.} = \frac{0.6745\,(1-0.73^2)}{\sqrt{22}} = \frac{0.308}{4.69} = 0.06$$

$$r = 0.73 + 0.06$$

As 'r' is more than six times the size of the probable error and as it is more than 0.5 there is a fairly high degree of positive correlation.

Table—7 **Pearson's Coefficient of Correlation applied to marks of Twelve Students.**

Student	Marks in English	Deviations from the mean x	x^2	Marks in Maths.	Deviations from the mean y	y^2	xy
A	75	22	484	30	−18	324	−396
B	40	−13	169	55	7	49	−91
C	30	−23	529	68	20	400	−460
D	85	32	1024	18	−30	900	−960
E	33	−20	400	70	22	484	−440
F	90	37	1369	15	−33	1089	−1221
G	25	−28	784	80	32	1024	−896
H	10	−43	1849	85	37	1369	−1591
I	95	42	1764	10	−38	1444	−1596
J	65	12	144	35	−13	169	−156
K	55	2	4	45	−3	9	−6
L	33	−20	400	65	17	289	−340
Total Average	636 53		8920	576 48		7550	−8153

$$Sx = \sqrt{\frac{8920}{12}} = \sqrt{743.3} = 27.3$$

$$Sy = \sqrt{\frac{7550}{12}} = \sqrt{629.1} = 25.1$$

$$r = \frac{-8153}{12 \times 27.3 \times 25.1} = \frac{-8153}{8222.8} = -0.99$$

$$P.\ E. = \frac{0.6745\ (1-0.99^2)}{\sqrt{12}} = \frac{0.013}{3.46} = 0.004$$

$$r = -0.99 \pm 0.004$$

As the coefficient is negative and is almost approaching unity the correction is almost perfect but negative. This the corroborated by the result already obtained by the graphic method (vide Fig. 3.)

Short-cut Method of Calculation:—In the above group the number of observations is small but if it is large it becomes necessary to group them and the correlation table will then take the form of frequency distribution in two dimensions. If the deviations in each case are taken from the arithmetic average there will occur inconvenient fractions and elaborate calculations will have to be made. To get over this difficulty a convenient figure is *assumed* as the average and the deviations are then calculated in each case from this assumed average and not from the actual arithmetic mean. The standard deviation can be computed in such a case as already explained by using the formula :

$$S.\ D. = \sqrt{\frac{\Sigma fd^2}{N} - C^2} \times i$$

and the formula for computing the coefficient of correlation is

$$r = \frac{\dfrac{\Sigma xy}{N} - Cx \times Cy}{Sx \times Sy}$$

where C is the difference between the actual average and the assumed one i. e. $\frac{\Sigma fd.}{N}$ which is the algebraic total of the 'fd' column divided by the number of observations.

In using this formula it is better to neglect class intervals in all the cases i. e. in the computation of xy, Cx, Cy and the standard deviation with a view to reducing calculating as the class interval does form a common factor in the case of the numerator as well as the denominator.

Grouped Data :--For computing the Pearson's coefficient of correlation it will have first to be arranged as shown in the following example of ages of husbands and wives and standard deviations calculated in both the cases by taking assumed averages. For compiling the 'xy' column the values of x based on their deviations will have to be carefully calculated. For instance, there are 14 couples with husbands having the ages between 20 and 25 but wives are divided in two age groups (x) five in 15–20 and nine in 20–25. Their respective deviations will have to be given proper importance and so for computing xy we cannot take 14 as the deviations in the case of x but the sum of the various frequencies in age groups multiplied by their respective deviations shown below will have to be taken.

$$\begin{array}{rr}
9 \times (-3) = & -27 \\
5 \times (-2) = & -10 \\
\hline
 & -37 \\
\hline
18 \times (-2) = & -36 \\
10 \times (-1) = & -10 \\
\hline
 & -46
\end{array}$$

These figures will be entered in the 'Deviations Σx column and the values of xy will then be computed by multiplying the figures in Σx column by the figures in 'y' column.

Table—8. **Calculation of the Coefficient of Correlation of a Grouped Data**

Ages of husbands y		Ages of wives (x)									Deviations from the assumed mean y	fy	fy^2	Deviation Σx	Σxy
		15–20	20–25	25–30	30–35	35–40	40–45	45–50	50–55	Total f					
	mid-values	17.5	22.5	27.5	32.5	37.5	42.5	47.5	52.5						
20–25	22.5	9	5							14	–3	–42	126	–37	111
25–30	27.5		18	10						28	–2	–56	112	–46	92
30–35	32.5		7	15	11					33	–1	–33	33	–29	29
35–40	37.5			11	15	4				30	0	0	0	– 7	0
40–45	42.5				8	10	2			20	1	20	20	14	14
45–50	47.5					7	6	2		15	2	30	60	25	50
50–55	52.5				4		5	3	1	13	3	39	117	23	69
55–60	57.5					4			3	7	4	28	112	16	64
Total f		9	30	36	38	25	13	5	4	160		–14	580		429
Deviations from the assumed mean															
x		–3	–2	– 1	0	1	2	3	4						
fx		–27	–60	–36	0	25	26	15	16	–41					
fx^2		81	120	36	0	25	52	45	64	423					

Standard deviation $x = \sqrt{\frac{423}{160} - \left(\frac{-41}{160}\right)^2} \times i$

$$= \sqrt{2.64 - 0.06} \times i$$

$$= 1.6 \times i$$

Standard deviation $y = \sqrt{\frac{580}{160} - \left(\frac{-14}{160}\right)^2} \times i$

$$= \sqrt{3.62 - 0.1} \times i$$

$$= 1.9 \times i$$

$$r = \frac{\frac{\Sigma xy}{N} - Cx \times Cy}{Sx \times Sy}$$

$$= \frac{\frac{429}{160} - \left(\frac{-41}{160}\right)\left(\frac{-14}{160}\right)}{1.6 \times 19}$$

$$= \frac{2.68 - 0.02}{3.04}$$

$$= 0.87$$

Probable Error $= \frac{0.6745\,(1-0.87^2)}{\sqrt{160}} = 0.01$

Coefficient of Correlation = 0.87 ± 0.01

There is, therefore, a decided evidence of direct correlation.

Concurrent Deviation Method:—This method of computing a coefficient of concurrent deviations takes into account the deviations not from the arithmetic mean or from the moving average but from the preceding item and it considers only the direction of the deviation and not its extent. This gives unnecessary importance to trivial changes which sometimes may be due to insignificant factors. This can be overcome by taking a fairly large number cases for consideration.

In both the series the fluctuations are marked for each item noting against it (+) in the case of increase and (—) in the case of decrease. In case there is no rise or fall (=) is written against the item. All the pairs are then carefully observed and those having the same signs for both the items of the pair, known as the Concurrent Deviations, are marked plus in the column headed concurrences. If otherwise, a minus sign is entered in the next column bearing the title of disagreements. The concurrences are then totalled and the coefficient of concurrent deviations is computed by the use of the following formula:—

$r = \pm\sqrt{\pm\dfrac{2C-N}{N}}$ where C stands for the number of concurrent deviations and 'N' is one less than the number of pairs of items which is equal to the number of deviations measured.

In this formula it will be noticed that ± signs are at two places. If the value of $\dfrac{2C-N}{N}$ is negative it will not be possible to extract the square root of the negative quantity and in that case the minus sign will be introduced before the quantity and also before the radical. This will make it possible to extract the square root and will also retain the same sign as was in the original quantity.

Table-9. **Correlation by Concurrent Deviation method.**

Weeks	Average number employed daily		Average number of bales consumed daily		Concurrences	Disagreements
	Number in thousands	Deviations from the preceding week	Number in lakhs	Deviations from the preceding week		
I	384		21			
II	385	+	24	+	+	
III	362	—	20	—	+	
IV	348	—	22	+		—
V	384	+	26	+	+	
VI	395	+	26	=		—
VII	403	+	29	+	+	
VIII	400	—	28	—	+	
IX	385	—	27	—	+	
X	415	+	31	+	+	
XI	418	+	32	+	+	
					8	2

$$r = \pm \sqrt{\pm \frac{2C-N}{N}}$$

$$= \pm \sqrt{\pm \frac{2\times 8-10}{10}}$$

$$= \sqrt{\frac{16-10}{10}}$$

$$= \sqrt{\frac{6}{10}}$$

$$= 0.77$$

The result obtained is not very much different from the one obtained by using the Pearson Method (*vide* Table—4)

This method of concurrent deviations is very easy as compared with the Pearson's Method and is very useful in cases where we are interested in knowing only if a charge in one is accompanied by a change in the other either in the same or opposite direction. It is also suitable for the study of short-time changes but not in the case of long-time fluctuations.

Spearman's Method of Ranks:—This method of ascertaining the coefficient of correlation by ranks usually shown by the symbol p, the Greek 'rho' was invented by Professor Charles Sperman and is so called as it is based on the ranks of the variables. In each case they are arranged according to their sizes and ranks are assigned. The difference 'd' in ranks in the two cases of each variables is then ascertained and the coefficient of correlation by ranks is computed by using the following formula :–

$$p = 1 - \frac{6\Sigma d^2}{N(N^2-1)}$$

wherein Σd^2 is the sum of the squares of the differences in rank and N, the number.

The value of the Pearson's coefficient of correlation computed from the sum–product method is not usually the same as that of the coefficient of correlation ascertained by the Ranking Method and Professor Thurstone in his ***Fundamentals of Statistics*** has given the following table showing the value of Persons' Coefficient of correlation corresponding to various values of the Rank correlation coefficient.

Table—10 Correlation between Rank and Pearson's Coefficient

Rank correlation coefficient	Pearson's coefficient.
.00	.000
.05	.052
.10	.105
.15	.157
.20	.210
.25	.261
.30	.313
.35	.364
.40	.416
.45	.467
.50	.518
.55	.568
.60	.618
.65	.668
.70	.717
.75	.765
.80	.813
.85	.861
.90	.908
.95	.954
1.00	1.000

Ties in Ranks:—While assigning ranks sometimes there occur two or more variations having the same value and in such a case one of the following methods, preferably the second, is made use of.

(a) **The Bracket Rank Method:**—All such variates are given the same rank that is the one in which the tie occurred and the next variate is given the rank which it would have been given had there been no such tie.

(b) **The Mid-Rank Method:**—In this case all such variates are given the same rank but it is not the one in which the tie occurred but the average of the ranks of places occupied by all such variates. The next variate then gets as usual the rank which it would have had it not been for such a tie. It will be seen from the following example that while assigning ranks in the case of marks in English 'E' and 'L' have been given 8.5 rank

i. e. the average of 8 and 9 and the next rank given to C is 10. In the bracket rank method both E and L would have got 8th rank while C would have got the same rank i. e. 10th.

Table–11. Coefficient of Correlation by Ranking Method (Marks of twelve students as per Table—3.)

Students	Marks in English x	Marks in Maths. y	Their ranks in x	Their ranks in y	Difference between ranks d	Difference squared d^2
A	75	30	4	9	–5	25
B	40	55	7	6	1	1
C	30	68	10	4	6	36
D	85	18	3	10	–7	49
E	33	70	8.5	3	5.5	30.25
F	90	15	2	11	–9	81
G	25	80	11	2	9	81
H	10	85	12	1	11	121
I	95	10	1	12	–11	121
J	65	35	5	8	–3	9
K	55	45	6	7	–1	1
L	33	65	8.5	5	3.5	12.25
Total						567.5

$$\text{Coefficient of correlation} = 1 - \frac{6 \times 567.5}{12(144-1)}$$

$$= 1 - \frac{3405}{1716}$$

$$= -\frac{1689}{1716}$$

$$= -0.98$$

By transposing this from the table of equivalent values (Table no. 10) we get approximately—0.99 as equivalent value for 'r' which is actually the value as can be ascertained from the example already worked out.

The coefficient of correlation by the Pearson's Method can be computed only if the original values of the variables are known. But by the Ranking Method it can be calculated even if only the ranks of the variables are known. Again the rank coefficient can be computed with greater ease than is possible in the case of the Pearson's coefficient. This method of ranks can be profitably used in cases where the number does not exceed thirty.

Spearman's Foot-rule:—Spearman has devised another formula which is easier to compute but it gives only a very rough estimate of the measure of correlation and hence it is not usually used.

$$R = 1 - \frac{6\Sigma g}{N^2-1}$$

In this formula Σg stands for the total of gains in ranks or in other words positive differences in rank as shown in the column 'Difference between ranks' in the previous example. As said already, the result obtained by this formula will give only a rough estimate.

Ratio of Variation:—In case of correlation we compare the deviations from the type to see if the fluctuations are direct or inverse associated and also to ascertain the degree of this association. It often happens that the variables fluctuate together directly or inversely and yet the deviations are not absolutely or proportionately the same. The price of an article may rise with a rise in demand for it but its absolute or proportionate rise may not be the same. It is important to know how much change in one variable, on the average, results from a given change in the other. Will 50% rise in the demand lead to 50% rise in the price of the article? When the dearness allowance of twenty percent is paid to workers, is their expenditure on food going up by twenty percent? The ratio of variation, which is taken as the average ratio between the proportional or percentage deviation of the two curves from the selected base, is a corollary of the coefficient of correlation and is often confused with it because of the close relation of the one with the other.

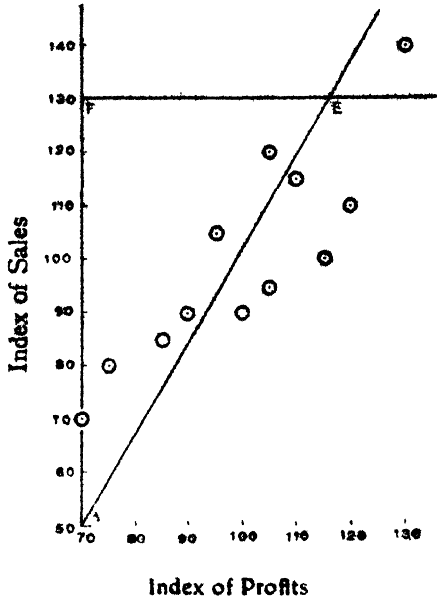
The Galton Graph:—The easy method of ascertaining this ratio of variation is by plotting the corresponding figures in the subject and the relative, the variables having larger proportional deviations forming the subject and the other, relative. But with a view to making allowance for the difference in variability of the subject and the relative and to obtain the percentage variation in the latter, for a given percentage variation in the former, index numbers are calculated in both in cases based on the arithmetic mean in each case.

The Regression Line:—When these points are plotted they will be too much scattered where there is an absence or correlation and in that case the problem of the ratio of variation does not arise. When there is perfect correlation, these plotted points will form more or less a straight line which will point towards the lower left corner in the case of positive correlation and towards the lower right corner when there is negative correlation. Such a case is rarely met with but when there is partial correlation, the points are very often found to form a bend. A straight line should then be drawn keeping the trend of the points and having as far as possible equal number on either side of it and it should be the best possible fit to the points in question. It is no doubt a difficult job and for the same set of points different persons are likely to draw *slightly* different lines. This can be better done by taking a piece of thread, and then trying different positions of a straight line till the one which will be the best fit is fixed up. This line will correspond to a line of trend describing the average change and is known as a Line of Regression. The use of the word 'Regression' is traced to the early studies of the relation of the heights of fathers and sons by Francis Galton as early as the second half of the nineteenth century. The heights of sons appeared to regress as compared to those of the fathers and so he called the line which showed graphically the average relation as the line of regression.

Table—12. Ratio of Variation between Sales and Profits

Year	Sales		Profits	
	In lakhs of of rupees	Index	in lakhs of rupees	Index
1935	72	120	42	105
1936	84	140	52	130
1937	66	110	48	120
1938	60	100	46	115
1939	48	80	30	75
1940	42	70	28	70
1941	54	90	36	90
1942	63	105	38	95
1943	51	85	34	85
1944	57	95	42	105
1945	69	115	44	110
1946	54	90	40	100
Total	720	1200	480	1200
Average	60	100	40	100

The ascertain the ratio of variation between the production and profits of the given historical series the Galton graph is constructed according to the usual practice of taking the ordinate to represent the subject and the abscissa the relative.



Variation between Sales and Profits

Fig. 5. Galton Graph

The straight line AB which corresponds to the line of trend is the regression line and meets the ordinate at A. If the variations in the subject and in the relative are of equal proportion their points when plotted will form the straight line which will make an angle of 45° with the abscissa as well as with the ordinate. This line is known as the Line of Equal Varitation. For measuring the ratio of variation a straight line parallel to the base is drawn from any point in the ordinate to meet the line of regression at E and the ordinate at F. The ratio of the average variations of the relative to the average variations of the subject will be $\frac{EF}{AF.}$ As EF and AF in the above case come to 23 and 40 units the ratio of variation comes to 0.575. This shows that for every change of one percent in the subject there will be approximately a change of 0.575 percent in the relative. The ratio of regression is the complement of this portion and therefore, comes to 0.425.

Exercises

1. Define correlation and distinguish between positive and negative correlation.

2. Give the important methods of ascertaining correlation and state with reasons the method you recommend.

3. Write short notes on :—

   Subject and relative, lag and lead, probable error, Spearman's Foot Rule and Scatter Diagram.

4. Describe the Pearson's method of measuring correlation and give rules for interpreting the coefficient of correlation.

5. Explain any two methods of the following —

   (a) Concurrent Deviation Method.
   (b) The Bracket Rank Method.
   (c) Spearman's Method of Rank.
   (d) The Graphic Method.
   (e) The Mid-Rank Method.

6. What is meant by a Regression Line ? Describe in detail the method of drawing it

7. What is a Lag? How will you trace it ?

8. In the following table are given figures in crores of rupees of Clearing Houses in Calcutta and Bombay. Calculate the coefficient of correlation.

	Calcutta	Bombay		Calcutta	Bombay
1925	1018	519	1931	756	640
1926	959	421	1932	747	646
1927	1024	398	1933	824	646
1928	1088	543	1934	864	683
1929	998	800	1935	939	750
1930	893	712	1936	899	721
			1937	993	837

9. What is correlation? Calculate the coefficient of correlation from the following table:—

Year		Area under tea in '000 acres	Production in '000,000 lb.
1925	...	728	354
1926	...	739	393
1927	...	756	391
1928	...	776	404
1929	...	788	433
1930	...	804	391
1931	...	807	394
1932	...	809	434
1933	...	818	384
1934	...	826	399
1935	...	832	394
1936	...	834	395

(1939–Oct.)

10. Calculate the coefficient of correlation from the following figures. Is there any marked correlation between the production and price of tea?

Year	Production in crores of lb.	Price per lb.	Year	Production in crores of lb.	Price per lb.
1920–21	34	0—8–6	1925–26	36	1—0–6
1921–22	27	0—8–0	1926–27	39	0–15–0
1922–23	31	0–11–0	1927–28	39	0—9–6
1923–24	38	0–15–0	1928–29	40	0–13–0
1924–25	38	0–15–0	1929–30	44	0–12–0

11. **From the data given calculate the coefficient of correlation between the area under cotton and its price as also between the price of cotton and its yield.**

Year	Area under Cotton in lakhs of acres	Price per candy	Yield of lakhs of bales of 400 lb.
1914–15	152	193	49
1915–16	114	300	51
1916–17	138	414	50
1917–18	154	594	45
1918–19	144	676	40
1919–20	153	549	53
1920–21	141	320	49
1921–22	117	483	60
1922–23	136	481	63
1923–24	154	659	60

12. What is coefficient of correlation ? Find its value for the following pairs of values.

X	Y	X	Y
3	11	10	6
6	9	9	6
6	4	12	5
12	8	14	3
9	7	17	0
12	7		

(1947–Oct.)

13. Calculate the correlation coefficient between X and Y given by the following table.

X	Y
25	52
29	57
38	69
43	76
45	79
49	84
55	92
59	96

(1949–Oct.)

14. **What is meant by a correlation coefficient? What is its usefulness? The following table gives the income and expenditure on food of 10 working class families in Ahmedabad in 1932–33.**

Income of family in rupees	Expenditure on food in rupees
20	10.2
25	12.3
35	15.9
45	19.6
55	22.6
65	26.8
75	29.4
85	32.0
100	42.5
105	43.0

Calculate the correlation coefficient between income and expenditure on food.

(1949–March).

15. Calculate the coefficient of correlation from the following table and explain the meaning of probable error.

Year	Acreage under cotton in millions	Mean annual price per candy
1927–28	25	347
1928–29	27	367
1929–30	26	341
1930–31	24	232
1931–32	24	184
1932–33	22	201
1933–34	24	198
1934–35	24	208
1935–36	26	225
1936–37	25	210

(1940-Mar.)

16. **Calculate the correlation coefficient between (i) estimated yield per acre and (ii) the average rates of assessment per**

acre in the Settlement Groups of a district from the following data :—

Estimated yield per acre in Maunds	Average assessment per acre in Annas	Estimated yield per acre in Maunds	Average assessment per acre in Annas
2.7	18	4.2	26
3.5	21	3.5	19
2.1	9	3.9	30
5.3	34	4.5	33
2.9	19	6.1	40
3.6	21	6.6	40
5.0	29	5.4	27
		4.4	13

17. Ten students got the following percentage of marks in Mathematics and Statistics :

Student	1	2	3	4	5	6	7	8	9	10
Marks in Mathematics	78,	36,	98,	25,	75,	82,	90,	62,	65,	39.
Marks in statistics	84,	51,	91,	60,	68,	62,	86,	58,	53,	47.

Draw a Scatter Diagram and find the coefficient of correlations.

(1946–Oct.)

18. The following table gives the marks obtained in two subjects by students in a class.

Determine the extent to which capacity in one subject is correlated with capacity in the other.

Serial number of students	Marks in Subject A	Marks in Subject B
1	81	71
2	71	60
3	65	47
4	56	64
5	47	73
6	83	53
7	72	61
8	67	77
9	60	82
10	47	42
11	86	64
12	75	69
13	69	51
14	59	51
15	39	61

(1944–Oct.)

19. (*a*) Draw a Scatter Diagram to represent the following data sharing for apple tress the size of crop (x) and the percentage of worm-eaten apples (y) on each tree.

X 8, 6, 11, 22, 14, 17, 18, 24, 19, 23, 26, 40

(hundrdes of apples)

Y 59, 58, 56, 53, 50, 45, 43, 42, 39, 38, 30, 27.
With the help of the Scatter Diagram estimate roughly the number of worm-free apples for a tree with a yield of 2000 apples.

(*b*) Calculate the coefficient of correlation between size of crop and percentage of worm-eaten apples from the above data. (1946–Mar.)

20. Find the coefficient of correlation between marks obtained by candidates at an examination in two subjects A and B from the following data :—

Subject A Max. marks 50	Subject B. Max. marks 50.					
	11–15	16–20	21–25	26–30	31–35	Total
1– 5					1	1
6–10	1	1	8	7	1	18
11–15	1	2	4	14	4	25
16–20			7	13	6	26
21–25			2	4	1	7
26–30			1			1
31–35				1		1
Total	2	3	22	39	13	79

21. **What is correlation? Find the coefficient of correlation between the heights and weights of 90 students as given below.**

Height in inches.	Weights in lb. 90–	95–	100–	105–	110–	115–	120–	125–	Total
54—	1	1	1						3
56—		2	3	2					7
58—		1	4	6	1				12
60—			5	10	6	2			23
62—				9	8	6	2		25
64—				4	3	5	1		13
66—					3		1	1	5
68—						1		1	2
Total	1	4	13	31	21	14	4	2	90

22. Calculate the coefficient of correlation between the ages of husbands and wives from the following data.

Husbands	Wives 15–	20–	25–	30–	35–	40–	45–	50–	Total
55—60					4			3	7
50—55				4		5	3	1	13
45—50					7	6	2		15
40—45				8	10	2			20
35—40			11	15	4				30
30—35		7	15	11					33
25—30		18	10						28
20—25	9	5							14
Total	9	30	36	38	25	13	5	4	160

23. Calculate coefficient of correlation between the price and the spinning value from the correlation table given below for 47 different varieties of cotton :—

Spinning Value of Cotton

Adjusted price in rupees per Candy	12	17	22	27	32	37	42	Total
310							1	1
290					1	1		2
270			1	1	1			3
250		1	4	6	3	1		15
230		1	2	4	6			13
210	2			2		1		5
190	1	2	1			1		5
170	2	1						3
Total	5	5	8	13	11	4	4	47

(1941–Mar.)

24. Explain how a correlation table is prepared and draw up a brief report on the relation between the size of the farms and the net profits from the following data.

Size of Farm Acres ——> Net Profits ↓ Rs.	0 to below 5	5 ,, ,, 7.5	7.5 ,, ,, 10	10 ,, ,, 12.5	12.5 ,, ,, 15	15 ,, ,, 17.5	17.5 ,, ,, 20	20 ,, ,, 22.5	22.5 ,, ,, 25	25 ,, ,, 27.5	27.5 ,, ,, 30	30 and above	Total
100 and above	..	2	1	.	2	1	4	.	...	1	1	5	17
50 to below 100	2	1	2	4	3	3	1	.	.		3	2	19
0 ,, ,, 50	2	4	7	6	5	2	6	2	1	2	2	1	40
–50 ,, ,, 0	18	11	12	14	11	7	6	5	1	..	1	2	88
–100 ,, ,, –50	30	27	25	15	10	10	4	4	.	.	...	1	126
–150 ,, ,, –100	11	23	24	16	18	6	3	3	2	1	1	1	109
–200 ,, ,, –150	5	8	9	21	22	6	2	3	.	2	..	3	81
–300 ,, ,, –200	.	4	5	11	12	7	4	3	3	1	3	5	58
below–300	...	..	...	2	2	3	1	...	3	2		6	19
Total	68	80	85	89	85	45	31	20	10	9	9	26	557

(1943–Oct.)

25. **Calculate the coefficient of correlation from the following data :—**

Yield of wheat per acre—in bushels	Nitrogen applied in pounds per acre									
	0–20	20–40	40–60	60–80	80–100	100–120	120–140	140–160	160–180	Total
32–36		...	...	5	16	12	4	5	2	44
28–32	...	...	1	20	21	8	4	1	...	55
24–28	..	..	16	19	..		..	.	..	35
20–24	..	...	13	...	..	..	...	...	...	13
16–20	...	12	...	...	...	..	...	...	..	12
12–16	...	8	...	...	...	...	...	...	...	8
8–12	3	5	...	...	...		...	..	...	8
4–8	10	...	...	...	...	...	...	...	...	10
0–4	8	...	...	...	...	..	...	...	...	8
Total.	21	25	30	44	37	20	8	6	2	193

(1939–Mar.)

26. Calculate the coefficient of correlation between the marks obtained by 80 students in the Terminal (X) and Annual (Y) Examinations in "Trade and Statistics".

X Y	21–30	31–40	41–50	51–60	61–70	71–80	81–90	Total
21–30	4							4
31–40	5	3	11					19
41–50		2	10	8				20
51–60		6	10	5				21
61–70				4	3			7
71–80					2	2	1	5
81–90						3	1	4
Total	9	11	31	17	5	5	2	80

(1947–Mar.)

27. Calculate the coefficient of correlation between x and y from following data.

x y	10–20	20–30	30–40	40–50	50–60	60–70	70–80	Total
10–20				2	2	1		5
20–30			4	9	7	3	2	25
30–40		3	7	14	10	8	4	46
40–50	1	6	4	4	1			16
50–60	3	4	1					8
Total	4	13	16	29	20	12	6	100

CHAPTER XIII

INTERPOLATION AND EXTRAPOLATION

In the course of a statistical study it very often transpires that one or more figures in a statistical series are anyhow missing and they are required to be inserted by estimating them with as much care and accuracy as is possible. It is likely that while compiling the table of sales of an establishment for the last thirty years, the sales figure for a particular year is not available due to the destruction of records by fire or because of some other reason. We might have to ascertain whether there has been a correlation between the sales of this establishment and those of the other. If this missing amount is not inserted the coefficient of correlation is difficult to be ascertained. A teacher may be interested in the quarterly results of his students and the record may be incomplete because the test was not taken for a certain quarter in the period. A census of a country is normally taken once in ten years but birth and death rates are compiled for each year and are expressed in relation to 1000 persons of a place. A census is usually a decennial process and as it involves a considerable expenditure as well as energy actual figures of population of each year cannot be had and to calculate the birth and death rates on the preceding census figures will be wrong as the population of a place does not remain stationary. It will be, therefore, necessary to have intercensual figures for this purpose. Again, sometimes it is found that the data pertaining to a particular topic is grouped by different persons in different types of groups and this makes it almost impossible to make a comparison. For example, a certain municipality may group its population by making age-groups of fifteen years while the other may have them of ten years. In such a case it is not possible to calculate the corrected death rate of that place to effect a proper comparison of the death rates of the two places under consideration. The problems of the type described above can be, to a great extent, solved by interpolation.

The dictionary meaning of the word 'interpolate' is 'to insert as a spurious word or passage in a manuscript or book or to corrupt or vitiate by the insertion of the new matter.' The word 'Interpolation,' therefore, has come to mean a spurious word or passage inserted but in mathematics as well as in statistics this word is not used in such a bad sense. On the contrary, in conveys the idea of filling up intermediate terms of a series according to the law of the series Interpolation is, therefore, the insertion of the most likely estimate based on certain assumptions such as, the prevalence of the normal circumstances and possession of the knowlege of the course of events by the interpolator. If, for example, the missing sales figure for a certain year is to be interpolated or inserted in the table of sales for the last thirty years it will be of great importance to know if in that particular year normal circumstances prevailed or not. It is likely that there may have been a violent outburst of a disease like plague or influenza because of which sales might have been considerably reduced in that particular town. It is equally possible that the year may have proved to be an exceedingly good year from the sales point of view. In normal times normal results are anticipated and so the estimate prepared for an average year will not be suitable for an abnormal one. Again, the course of events requires to be given proper importance. The article in question might have been in fashion at one time but not at the other or other substitutes of even better type might have been available thus adversely affecting the sales.

Interpolation supplies us with the missing link while extrapolation also called 'Projection' helps us in forecasting. A sales manager has to order his stock of goods in advance and to keep an adequate stock of goods particularly of seasonal goods. Woollen goods in India can be sold only in certain months of the year as such goods will be in demand only for winter wear which is required for a few months only. If the stock falls short of demand there may not be enough time left to get the necessary supply in time. If correct forecasting is not made and too large a stock is purchased the capital may remain locked in goods which cannot be sold for a considerable period of time i. e. up to the coming of the next season. In the modern times when there is production on a very large scale extrapolation plays a very important part. It represents a logical guess based, as in the case of interpolation, on the existence of normal circumtances and adequate knowledge of the course of events.

Assumptions:—The degree of accuracy in case of interpolation as well as extrapolation will mainly depend upon the conformity of the assumptions to the actual circumstances. It is generally assumed that there are no sudden or fortuitous changes from one period to the other and that the rise or fall in the absence of evidence to the contrary has been more or less uniform. When we are to forecast sales of the coming year we naturally assume that the trend of sales will remain the same and so the rise or fall will be maintained and that there will be no possibility of any sudden change in the situation. But if the price of the article under consideration is, all of a sudden, controlled by the Government or it is decided to ration it out and not to issue it as freely as is usually done, our forecast is sure to be upset. Besides, there are certain cases where psychological influences upset the estimates. The sales of foreign goods in India after the Swadeshi movement went down considerably. Quotations in the share market may have been rising steadily but all of a sudden due to some psychological influence the rates go down. In short it is on these assumptions that the correctness or otherwise of interpolations as well as forecasting largely depends.

As all the figures that are interpolated or extrapolated are only estimates made under certain assumptions and not the real figures it is necessary that they should be indicated as such. In the absence of this there is a danger that some one may take them as actuals and not as estimates and base his estimates on these estimated figures and, consequently, his results are likely to go off the mark. In matters of crop estimates it happens that sometimes the required crop–returns for a certain district are not received in time and as the returns are to be submitted on the particular day estimated figures are put in with a note to that effect. Even then it is necessarry that the note should explain the method adopted in arriving at the required figure as it is likely that different methods may be made use of by different persons.

Methods:—The device commonly used is to put the previous year's figure unless the circumstances have considerably changed or to take an average of the last three or five years to remove the effect of fluctuations. The other method is to estimate the missing figure by ascertaining the ratio as illustrated in the following example of the wholesale prices of tea at important centres.

Table No. 1—**Wholesale prices of tea per lb.**

Centres	April 1935	April 1936
Assam	0–7–3	0—8—2
Cachar	0–6–4	0—8—0
Sylhet	0–6–6	0—8—1
Darjeeling	0–8–9	0–13—8
Dooars	0–7–5	0—8–10
Tarai	0–7–0	0—8—3

Assuming that the price for the Cachar centre is not known for April 1936, the average of the prices for the remaining centres will be ascertained for both the years. For April 1935 the average comes to

$$\frac{0\text{–}7\text{–}3 + 0\text{–}6\text{–}6 + 0\text{–}8\text{–}9 + 0\text{–}7\text{–}5 + 0\text{–}7\text{–}0}{5} = 0\text{–}7\text{–}4.6 \text{ While}$$

that for April 1936 it is

$$\frac{0\text{–}8\text{–}2 + 0\text{–}8\text{–}1 + 0\text{–}13\text{–}8 + 0\text{–}8\text{–}10 + 0\text{–}8\text{–}3}{5} = 0\text{–}9\text{–}4.8.$$

The rise, therefore, is from 0–7–4.6 to 0–9–4.8 i. e. in the ratio of 886 : 1128. The estimated price for the Cachar centre for April, 1936, by this method will be $\frac{1128 \times 76}{886} = 0\text{–}8\text{–}1.9$. The actual quotation is 0–8–0 and is not very far from the estimate.

The important methods of interpolation most commonly used are two, ***Graphic*** and the ***Algebraic.*** The former is suitable only in the case of a continuous series and in the case of problems of cyclic character it is found that the results obtained are fairly accurate under normal circumstances.

The Graphic Method:—When interpolation is to be done by the graphical method the available data of the series is represented graphically by using the abscissa to indicate the various intervals of time and then effort is made to complete the remaining portion of the curve by free hand or by joining the preceding and succeeding points which will give us the missing figure as will be seen from the following example.

Table No. 2—**Statistics of Indian Railways**

Year	Total mileage (in hundreds.)
1928	397
1929	409
1930	417
1931	422
1932	428
1933	429
1934	429
1935	430

Let us suppose that the mileage for the year 1931 is not available and it is required to be interpolated, it being, however, known that it was not an abnormal year,

On a piece of graph paper the available data will be represented by showing on the base line the various years beginning with 1928. The millage will be indicated by the ordinate by taking a convenient scale. From the diagram it will be seen that the points A, B and C are representing the mileage for the years 1928, 1929 and 1930. The mileage for the year 1931 is not known and that for the years 1932, 1933, 1934 and 1935 has been indicated by the points D, E, F and G. The easy method is, to complete the curve A G by joining C to D by *free-hand* cutting the ordinate for the year 1931 at L which gives nearly 423 as mileage for that year. The actual figure is 422 and the difference is only of one hundred miles which comes to about 0.2% The other method is of joining the two points by straight line cutting the said ordinate at M thus giving roughly 422.5 as the mileage. The actual difference in this case is of one-half unit and the percentage of error is almost 0.1%.

Interpolation of mileage in 1931

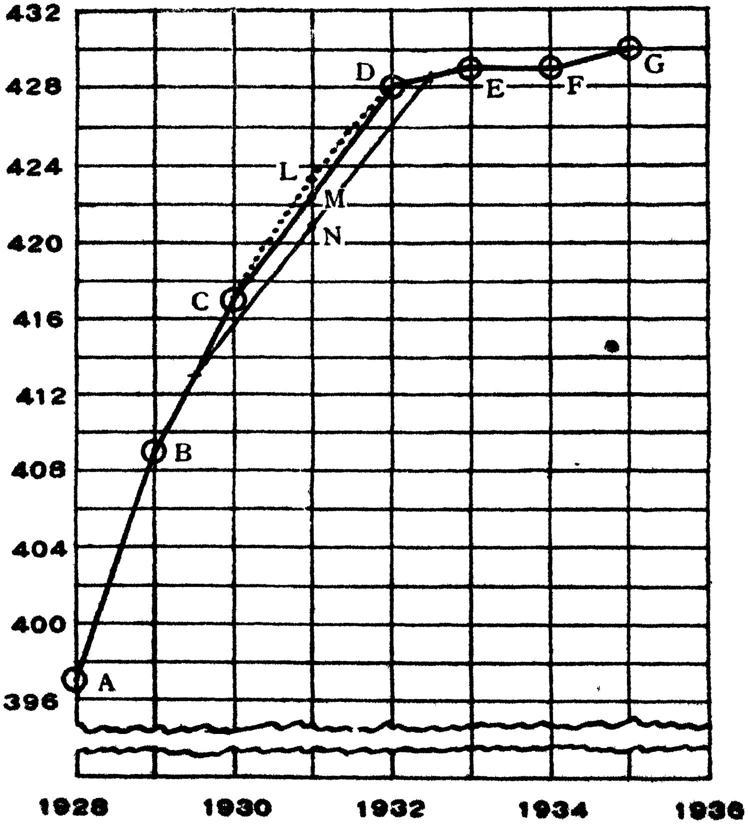


Fig. No. 1

In this case we have based our estimate only on the figures of the preceding and succeeding years and have interpolated the figure by joining the points. But when we have in our possession figures for other years as well it is better to take them into consideration, as this will help us to arrive at a more correct estimate based on the figures of more than two years. The average of the two preceding and the succeeding years should be obtained which can be done by bisecting the lines BC and DE at X and Y and then joining these points cutting the ordinate for the year 1931 at N giving 421.75 units as the mileage for the year and it is far more correct, the error being only 0.06 percent. In this method the figures for more than two years have been taken into consideration and so the result is more dependable than the one arrived at by the previous method of joining only the points of figures for two years only.

In the above case the figures for the various years except one were given and so the interpolation was found to be more correct. But when the actuals are given for larger intervals of time as in the case of a census the possibility of accuracy becomes less and the second method will in such a case give better results. The census is taken once in ten years and so the actuals will be available for every tenth year. The figures for the intermediate years will have to be obtained by interpolation. In such a case the population curve will be drawn with the help of the various census figures and for the intermediate or intercensual years the figures will be obtained by observing the values of the points where this curve will cut the ordinates of the years concerned.

Use of Correlation Curves:—If there exists a good degree of correlation between two series aud if a figure is to be interpolated in one of them it will be very helpful to study the close relation between them and then to complete the portion of the missing curve by maintaining the same degree of resemblance between them. This is a very simple process and will be found of great utility in statistical matters such as rainfall or temperature at two places, prices and circulation of money, production and prices, etc.

The Algebraic Method:—The graphic method is suitable for the series having cyclic fluctuations but in the case of those that do not show a fair degree of periodicity it fails to give a fairly accurate result. In such a case the algebraic method is found to be more suitable, assuming of course in both the cases that the changes as well as the rate of changes in the series are more or less continuous and not sudden. The following are some of the algebraic methods used for interpolation.

Fitting with a Parabolic Curve:—The first method consists of taking X and Y as variables and then assuming that the value of Y depends upon that of variable A in such a way that from the given value of X, that of Y can be estimated from the following eqyation of the parabolic curve which is assumed to pass through all the points representing the different values of Y.

$$Y = A + BX + CX^2 + DX \ldots\ldots\ldots NX_n \quad \ldots \quad \ldots \text{(i)}$$

Table No. 3—**Population of India**

Year	Population in millions
1881	254
1891	287
1901	294
1911	315
1921	319
1931	353
1941	389

In the above table decennial figures are given and it is required to interpolate the population for the year 1935.

If X stands for the year and Y for the population the equation (i) given already stands true for all values of X and Y.

The four known points would help us to determine the value of a, b, c, d.—

Taking $x = \frac{X-1931}{10}$

$\therefore \quad y = 353 - y$

the equation (i) then becomes $y = A + Bx + Cx^2 + Dx$ ... (ii)
when X is 1911 Y is 315

$\therefore \quad x = \frac{1911 - 1931}{10} = -2$

$y = 353 - 315 = 38$

Substituting these values in equation (ii) we get

$38 = A - 2B + 4C - 8\ D$ ... ... (iii)

when X is 1921 Y is 319

$\therefore \quad x = \frac{1921 - 1931}{10} = -1$ and $y = 353 - 319 = 34$

Substituting these values in equation (ii) we get

$34 = A - B + C - D$ ... ... (iv)

when X is 1931 Y is 353

$\therefore \quad x = \frac{1931 - 1931}{10} = 0$ and $y = 353 - 353 = 0$

Substituting these values in equation (ii) we get

$0 = A + o + o + o$ ... ... (v)

when X is 1941 Y is 389

$\therefore \quad x = \frac{1941 - 1931}{10} = 1$ and $y = 353 - 389 = -36$

Substituting these values in equation (ii) we get

$-36 = A + B + C + D$ ... ... (vi)

From the above equations we will have to calculate the values of A, B, C and D.

By taking equation (iv) and (vi) and adding them we get

$34 = A - B + C - D$ ... ... (iv)

$\underline{-36 = A + B + C + D}$ ... ... (vi)

$-2 = 2A + 2C$

or $-1 = A + C$ ... ... (vii)

As $A = 0$ ... ... (v)

we get $-1 = 0 + C$ or $C = -1$

By multiplying equation (iv) by 2 and then subtracting equation (iii) from it, we get

$$68=2A-2B+2C-2D \qquad \ldots \qquad \ldots \qquad \ldots\ \text{(iv)}\times 2$$

$$38=A-2B+4C-8D$$

$$\underline{-\qquad -\qquad +\qquad -\qquad +}$$

$$30=A \qquad -2C+6D$$

Substituting the values of A=0 and C=− 1 we get

$$30=0+2+6D$$

$$\therefore\ D=4\frac{2}{3}=\frac{14}{3}$$

Substituting these values an equation (iv) we get

$$34=0-B-1-4\frac{2}{3}$$

$$\therefore\ B=-39\frac{2}{3}=-\frac{119}{3}$$

Again by substituting the values of A, B, C. and D in (ii) we get

$$y=0-\frac{119}{3}x-x^2+\frac{14}{3}x^3 \qquad \ldots \qquad \ldots \qquad \ldots\ \text{(viii}$$

Now when X is 1935

$$x=\frac{1935-1931}{10}=\frac{2}{5}$$

Substituting this value of x in equation (viii) we get

$$y=0-\frac{119}{3}\times\frac{2}{5}-\frac{4}{25}+\frac{14}{3}\times\frac{8}{125}$$

$$=0+\frac{-5950-60+112}{375}$$

$$=-\frac{5888}{375}=-15.7$$

Substituting this value in equation

$y=353-Y$ we get

$$-15.7=353-Y$$

$$\therefore Y=368.7$$

By this method insurance premiums for the intervening period can be easily ascertained as will be seen from the following example.

The annual premiums to secure an endowment policy of Rs. 1000 with profits payable at the age of 60 or at death if it occurs earlier are as follows:—

Age at birth day	Annual Premium	
20	Rs. 23- 2	23.128
21	Rs. 22-13	23.81
22	Rs. 24- 9	24.56
23	Rs. 25- 6	25.38
24	?	
25	Rs. 27- 2	27.12

It is required to interpolate the missing figure of premium.

If X stands for the age at birthday and Y for the premium the equation

$$Y = A + Bx + Cx^2 + Dx^3 \quad \ldots \quad \ldots \quad \ldots \quad \text{(i)}$$

stands true for all values of X and Y.

Taking $x = X - 23$ $\therefore$ $y=25.4-Y$ the given equation then becomes·

$$y=A+Bx+Cx^2+Dx^3 \quad \ldots \quad \ldots \quad \ldots \quad \ldots \quad \text{(ii)}$$

When X is 21 Y is 23.8

$\therefore$ $x=21-23=-2$ and $y=25.4-23.8=1.6$

Substituting the values in equation (ii) we get

$$1.6=A-2B+4C-8D \quad \ldots \quad \ldots \quad \ldots \quad \ldots \quad \text{(iii)}$$

When X is 22 Y is 24.6

$\therefore$ $x=22-23=-1$ and $y=25.4-24.6=0.8$

By substitution we get $0.8 = A - B + C - D$ ... (iv)

Whed X is 23 Y is 25.4

∴ $x = 23 - 23 = 0$ and $Y = 25.4 - 25.4 = 0$.

∴ $0 = A$

When X is 25 Y is 27.1

∴ $x = 25 - 23 = 2$ and $y = 25.4 - 27.1 = -1.7$

By substitution we get $-1.7 = A + 2B + 4C + 8D$ (vi)

By adding equations (iii) and (vi) we get

$$\begin{array}{l} 1.6 = A - 2B + 4C - 8D \\ -1.7 = A + 2B + 4C + 8D \\ \hline -0.1 = 2A \quad + 8C \end{array} \quad \ldots \quad \text{(vii)}$$

But $A = 0$ ∴ $-0.1 = 0 + 8C$ ∴ $C = -.012$

Multiplying equation (iv) by 2 and adding it to (vi) we get

$$\begin{array}{l} 1.6 = 2A - 2B + 2C - 2D \\ -1.7 = A + 2B + 4C + 8D \\ \hline -0.1 = 3A \quad + 6C + 6D \end{array}$$

Substituting the values of $A = 0$ and $C = -.012$ we get

$-0.1 = 0 - .07 + 6D$ ∴ $D = -.005$

Substituting these values in equation (iv) we get

$0.8 = 0 - B - .012 + .005$ we get $B = -.807$

Again by substituting the values of A, B, C and D in equation (ii) we get

$y = 0 - .807x - .012x^2 - .005x^3 \ldots$ ... (viii)

when X is 24, $x = 24 - 23 = 1$.

Substituting this value in equation (viii) we get

$y = 0 - .807 - .012 - .005$

$= -.824$

Substituting this value in the equation $y = 25.4 - Y$

$-.824 = 25.4 - Y$ we get Y—26.224.

The premium, therefore, works to Rs. 26-3-7. The actual amount given in the Premium Table of the company in Rs. 26-4-0.

Lagrange's Formula—The other method is of using this formula. In that case the problem of interpolation of population for the year 1935 can be worked as under ;—

$$Yx = Y_o \frac{(x-x_1)(x-x_n)}{(x_o-x_1)(x_o-x_2)} \cdots \frac{\cdots (x-x_n)}{\cdots (x_n-x_n)} +$$

$$Y_1 \frac{(x-x_o)(x-x\ )\ \cdots}{(x_1-x\ )(x_1-x_2)\ \cdots} \cdots \frac{\cdots (x-x\ )}{\cdots (x_1-x_n)} +$$

$$\cdots \qquad \cdots \qquad \cdots \qquad +$$

$$Yn \frac{(x-x_o)(x-x_1)\cdots}{(x_n-x_o)(x_n-x_1)\ .} \cdots \frac{\cdots (x-x\ -_1)}{\cdots (x_n-x_n-_1)}$$

As it is required to estimate the population in 1935 we have

Year			Population	
1911	...	x	315	y
1921	...	x	319	y_1
1931	...	x_2	353	y
1941	...	x	389	y

$$\text{Y for } 1935 = 315\frac{(1935-1921)(1935-1931)(1935-1941)}{(1911-1921)(1911-1931)(1911-1941)} +$$

$$319\frac{(1935-1911)(1935-1931)(1935-1941)}{(1921-1911)(1921-1931)(1921-1941)} +$$

$$353\frac{(1935-1911)(1935-1921)(1935-1941)}{(1931-1911)(1931-1921)(1931-1941)} +$$

$$389\frac{(1935-1911)(1935-1921)(1935-1931)}{(1941-1911)(1935-1921)(1914-1931)} +$$

$$\therefore \text{ Y for } 1935 = 315\ \frac{14 \times 4 \times (-6)}{(-10)(-20)\ (-30} + 319\frac{24 \times 4 \times (-6)}{10(-10)(-20)} +$$

$$353\ \frac{24 \times 14 \times (-6)}{(20 \times 10 \times (-10)} + 389\ \frac{24 \times 14 \times 4}{30 \times 20 \times 10}$$

$$= \frac{315(-336)}{-6000} + \frac{319(-576)}{2000} + \frac{353(-2016)}{-2000} + \frac{389 \times 1344}{6000}$$

$$= \frac{315 \times 336 - 957 \times 576 + 1059 \times 2016 + 389 \times 1344}{6000}$$

$$= \frac{48}{600}\left\{\ 2205 - 11484 + 44478 + 10892\ \right\}$$

$$= \frac{46091}{125}$$

$$= 368.728$$

EXTRAPOLATION

Extrapolation or forecasting is the process of making definite estimates of future conditions by working on a systematic and reliable basis. It does not depend upon the application of some mysterious or magic process but upon practical analysis of the past and present conditions indicating the nature of probable future conditions. It is a guess based on the assumption that there would be no sudden changes and the conditions in the future will not change considerably.

Forecasting is subject to wider margin of error as compared to interpolation as the former relates to the distant future as the data to be obtained will not be as certain and accurate as in the case of the latter. There are possibilities of such changes as are not likely to be anticipated and this is very much true in economic statistics. Who could have forecast the economic disorder due to the transfer of population from Pakistan or Hindustan and vice versa after 15th August, 1947? All estimates of production and sales are upset because of this unexpected event.

The simple method of extrapolation is of continuing the curve of the time series of the observations recorded keeping in view the trend of the past events. If instead of using the natural scale graph paper the logarithmic graph paper is used or longarithms of the data are plotted on the ordinary graph paper the forecast is likely to be more correct because the *rate of change* in case of a phenomenon is more likely to be maintained than the absolute amount of change.

Population of India

Year	Population in lakhs	Logarithm
1891	2794	3.4462
1901	2839	3.4532
1911	3030	3.4814
1921	3057	3.4853
1931	3381	3.5290
1941	3889	3.5898

In case a forecast is to be made for estimating the population in 1951 the population figures given should be plotted and the curve obtained should be extended in the same strain cutting the ordinate for the year 1951. This point will give the estimated figure. The other and better way is to use logarithmic graph paper instead of ordinary scale graph paper for this purpose. The purpose can as well be served by plotting the logarithms of these figures on an ordinary graph paper. As said above due to sudden changes the forecast is not likely to be as correct, as it ought to have been, if circumstances would have remained normal in the country.

The work of forecasting is of a specialised nature and in foreign countries a number of institutions have been established to render forecasting services and of them Babson is of one of the oldest.

Exercise

1. What is interpolation ? What is its utility and how can it be done ?
2. On what assumptions does the accuracy of interpolation depend ?
3. Name the methods of interpolation and describe fully the one you recommend.
4. Write short notes on :—

   projection algebraic method of interpolation and Lagrange's formula.
5. Under what conditions extrapolation is likely to give correct results ? Explain the working of the graphic method for forecasting-
6. The Census figures for various years are given and you are required to interpolate the figure for 1921.

Years	Population in India (in millions).	Years	Population in India (in millions)
1881	253	1911	315
1991	287	1921	?
1901	294	1931	352

7. The following table gives the sales of a concern for the following years.

Years	Cloth in lakhs of yds	Years	Cloth in lakhs of yds
1885	125	1930	238
1920	163	1935	282
1925	204		

Assuming the conditions to be the same estimate the sales for the year 1940.

8. Estimate the population of Bengal in 1941 based on the figures given below :—

Year	Population in lakhs	Year	Population in lakhs.
1881	363	1911	455
1891	391	1921	467
1901	421	1931	501

9. The following table gives the number of factories in different years.

Year	No. of factories	Year	No. of factories
1898	1098	1918	3436
1902	1533	1922	5144
1906	1855	1926	7251
1910	2359	1930	8148
1914	2936	1934	8658

Estimate the number for 1938 by the Graphic method.

10. **Interpolate the missing figure by the graphic method.**

Statistics of sugar industry in India.

Year	No. of factories	Year	No. of factories
1929—30	27	1933—34	112
1930—31	29	1934—35	130
1931—32	32	1935—36	?
1932—33	57	1936—37	140
		1937—38	146

11. The number of students in a recognised institution of India for the last nine years is given below. Estimate the number for the year 1936–37.

Year	No. of students in lakhs	Year	No. of students in lakhs.
1927–28	111.6	1932–33	121.2
1928–29	115.5	1933–34	124.9
1929–30	119.0	1934–35	128 2
1930–31	120.6	1935–36	131.2
1931–32	121.2		

12. The following table gives premiums payable annually to secure an endowment policy of Rs. 1000 with profits payable at the age of 60 or at death if it occurs earlier.

Age at birth day	Annual Premium Rs.	Age at birth day	Annual Premium Rs.
20	23–2	25	27–2
21	23–13	26	?
22	24–9	27	29–1
23	25–6	28	30–2
24	26–4	29	31–5

Interpolate the amount payable at the age of 26.

CHAPTER XIV

INDEX NUMBER

While studying the different problems regarding prices, volume of trade, unemployment, deferred payment and others it is noticed that all the variables in the group do not register changes in the same direction. For instance, prices of some commodities will show rise while those of others may have gone down. Similarly imports of some articles might have increased showing at the same time a decrease in those of others. Under these circumtances, it becomes inconvenient to estimate the relative changes without making use of some ingenious device such as an an 'Index Number'.

An Index Number is a statistical device for estimating the relative movements of a statistical variable and for measuring changes in the magnitude of a phenomenon from time to time or even from place to place. Supposing we are interested in ascertaining whether the condition of agricultural labourer in 1946 with increased wages is better or worse than their life as it existed ten years ago i. e. 1936, it will be very difficult to do so. Prices of some articles might have gone up while those of others might have dropped down and those of some might have remained stationary. It will be possible for an expert to give a rough idea but it will be impossible to gauge the situation correctly. We shall, therefore, have to reduce these changes in the first place to a common denominator by turning them into percentages of similar measurements for the selected basic period and then to ascertain the average. This will give the index number or in other words a term in a time series expressed as a relative to a given base. If it is higher than 100 it can be concluded that prices have, on the whole, risen and then we will have to see if the rise in the wages is proportionate or not to express our opinion regarding their condition of life with the increased wages. The Index number is thus designed to show relative change or difference of a group of related variables.

W. S. Jevons is credited with the first attempt in this direction. In the latter half of the nineteenth century he made a careful study of prices and as early as 1865 published his study of prices with Index Numbers from 1782 to that date. Other topics such as wages, volume of trade, industrial activity could have as well been taken up but the problem of prices appears to have attracted the attention of workers in the field more than any other topic mainly because of the universal interest in this subject. Rich as well poor are interested in this problem as it ultimately touches their daily life. A consumer as well as a producer has to face this problem without exception. Again, it has wide applicability in all economic matters, may they relate to production, consumption, exchange or distribution. Besides this, the data required for the construction of Index Numbers can be had more easily from municipal and government offices as well as from the various chambers of commerce.

Construction of Index Numbers of Prices:—The general principles involved in the construction of Index Numbers for the various problems are the same but as those for prices are most commonly used it is proposed to outline at length the technique of their construction. Though this problem of constructing the Index Number appears to be simple it involves the consideration of the following major problems:—

1. **Purpose** :— Before actually constructing an Index Number it is very necessary to define very clearly its *purpose.* It may be that it is to be costructed for studying the condition of the labour class or it may be required for that of the upper class. It goes without saying that the requirements of these two classes are to a great extent different. Rise in prices of luxury articles only will have no effect on the budget of the poor class while it will affect the cost of living of the other class. It is, therefore, necessary to bear in mind the purpose while making the selection of articles and also of their qualities for the construction of the Index Number. Even the mode of living and costoms and usages of the people concerned will have to be taken into consideration at the time of constructing Index Numbers as they materially differ with different classes of people and due to inattention to these things, Index Numbers obtained will be far from satisfactory,

2. **Selection of India:** 'An Index Number of Prices shows the average percentage change of prices from one point of time to another and so it has to be based on the prices of articles ruling in the market at the particular time. It is not possible to take into account the prices of all the articles exchanged in that market and so it becomes necessary to select some important commodities the prices of which will help us in the determination of the Index Number. The selection of such articles is required to be done with great care. The first requisite in this case is that the articles selected should be those as are widely used by the class of persons for whom the Index Number is to be constructed. They should conform to the tastes and habits of the people. It is no use giving importance to articles like wheat ghee, silk, etc. when we are preparing an Index Number for labour. It is also necessary that the articles included in the list are such as can be easily distinguished and do not vary materially in quality.

The other point of importance is of fixing the number of items to be taken. Different persons have taken different numbers of items. Babson is satisfied with 10 items only while in the list to U. S. Bureau of Labour Statistics the number is as high as 450. The articles taken by Sauerbeck amount to 35 while 96 is the number fixed by Bradstreet. Even in the case of the Index Number of the two principal cities of India we find a great divergence. The Bombay wholesale Price Index Number is based on 40 items while that of Calcutta includes 72 items and for the same type of Index Number in Great Britain 150 items are taken. It is a matter of common knowledge that larger the number of items the lesser will be chance for error in the average obtained. We must, however, have a manageable number and also to aim at a reasonable standard of accuracy. To have a fair play of average the number should not be less than thirty and as far as possible not more than fifty. If the number is too large it will be difficult to collect reliable data and will also entail a number of calculations without proportionate rise in the accuracy of the results. What is needed to be seen most is not the number but representative character of the articles that figure in the list. Articles of food, clothing, rent, fuel and lighting, besides some miscellaneous items must find a place in the proper type of list.

3. **Prices:**—Closely related to the question of the selection of items is that of their price quotations which can be obtained from standard trade journals or from important dealers at representative centres. While compiling the price statements it should be born in mind that '*inverse* prices' should not be quoted i. e. the price given should not be so many units per rupee but so many rupees or annas per unit of a commodity. Coal should not be quoted as 16 seers to a rupee but as Rs. 2-8-0 per maund. Money prices vary inversely with quantity prices and so the quantity prices are more easy to work out in the construction of Index Numbers. Again, to ensure better results it is advisable to take a Standard Price which implies the price which can be regarded as representative of a given commodity for the whole interval under consideration. It would naturally be the average of prices during the interval under consideration. If weekly prices are to be taken it would be better to add up the quotations for all the days in the week and then to divide it by the number of quotations taken. For obtaining the monthly average it would be sufficient to take the average of weekly average prices and for yearly average, average of monthly quotations would serve the purpose. It would add to the utility of the Index Number if quotations are obtained from selected places and their average ascertained. This will not allow abnormal fluctuations to creep in due to changes of price factors in a place or two. Of course, in the days of controlled price system prevailing in the country the problem becomes a complex one as people do obtain their supplies at black market rates which are so difficult to register.

Whenever price quotations are obtained they should be for the same quality of the commodity as with the change in the quality there would be a difference in price. The other point to be noted is that wholesale prices should be preferred to the retail ones as the former fluctuate less as compared to the latter and also conform more to the general trend of ruling prices. Retail prices register the rise more quickly than the fall and are influenced more by local conditions. In India, there is no trained agency to collect and maintain price statistics and particularly in the villages and upcountry areas statistics collected are far from satisfactory. Great care and judgment is, therefore necessary in the task of compiling Index Numbers from this type of material.

4. **Weightage:**—The main purpose of an Index Number of prices is to ascertain the changes in the price level. In case of a simple average all commodities will be given equal importance. But in actual practice different commodities command a different degree of importance with different groups of people. It has been found from the study of working class budgets of the Bombay City that rice is consumed far more than Jowar but in Sholapur the position is exactly reverse. This differentiation is taken into account by giving proper weightage i. e. by giving more importance to certain items by assigning weights in proportion to their importance or by taking more quotations of different qualities of the same commodity,

5. **Choice of Base:**—As already stated an Index Number shows the average percentage of prices from one point of time to another and so it becomes necessary to reduce the average prices to relatives by selecting a suitable base which may be a *fixed base* or a *chain base.*

(a) **Fixed Base Method**:—In this method either the average price of some arbitrarily chosen year or the average of prices of a period of some years serves as the base for calculating the percentages. In taking the average price of a year there is the possibility of that particular year being an unusual year due to abnormal circumstances such as financial crisis, labour troubles or continuance of war. Averages based on such cases are bound to mislead us and to avoid this it is imperative to see that the period selected is normal. If otherwise, it is necessary to append a note explaining the peculiar circumstances. This difficulty can be got over by taking the average prices of a period of years which will help in reducing the effect of abnormalities to a very great extent. The period selected should be of at least five years if not more and then the prices of this period should be averaged to serve as base. If the data is to be considered at the end of the period it is much better to take as base the average of prices ruling during the entire period. This will be an ideal method but very often it is not easy to do so. In India the Wholesale price indices, All India Food, Industrial raw material, as well as, semi-manufactured indices are worked on a single-year-base.

It is customary to take 100 as the average price of the base year or period and then the price of each year is calculated as a percentage to the price of the base year or period. For instance, if the price of milk in the base year, 1939, happened to be annas six a seer and it is now nine annas the percentage price of milk based on the price of 1939 will be $\frac{9}{6} \times 100 = 150$

(b) **Chain Base Method**:—In the fixed base method the base once selected remains fixed and all percentages are based on the average price of the base year or period. In the Chain Base Method this is different. The base taken is not fixed and the relatives for each year are worked out on the prices of the preceding year. As in the case of chain there are rings concerned with one another by linking them with the preceding ones the relatives in this method are calculated upon the prices of the preceding year and the results are then linked together. It is mainly because of this chain-like process as will be seen from the example worked out that this method is known as *Chain Base Method.* The Chain Base Numbers are also called *Link Index Numbers.*

6, **Form of Average**:—For constructing an Index Number of Prices any form of average such as mean, median, mode, geometric and harmonic mean can be used. From the practical point of view median and mode are not found to be suitable because of their being erratic. The harmonic mean as well as the geometric mean are difficult to compute and so usually the arithmetic mean is employed in this process as it has the advantage of being easily intelligible and with proper system of weightage the results obtained are quite satisfactory.

Methods Adopted:—There are various methods of constructing an Index Number of Prices but all of them are not equally important.

(1) **Method of Aggregates**:—The working in this case is very simple and easy. Prices of different commodities per unit are taken for the years under consideration and their totals are then compared as shown below :—

Commodity	Unit	Average annual price	
		1926	1928
		Rs.	Rs.
Wheat ...	Per maund ...	5—8—0	5—6—0
Rice ...	,, ...	7—4—0	7—0—0
Dal ...	,, ...	5—12—0	6—2—0
Ghee ...	,, ...	64—0—0	60—0—0
Cloth ...	Per yard ...	0—8—0	0—10—0
Tea ...	Per pound ...	0—10—0	0—11—0
		83—10—0	79—13—0

The conclusion drawn will be that prices in 1926 were Rs. 83–10–0 while they have gone down to Rs. 79–13–0 in 1928. Bradstreet adopted this method. He devided 96 articles in 13 classes and then used to compare prices by this method of aggregates, In this method there will be no base year and the prices of any two years can be compared simply by taking totals of commodity prices, all items being given equal importance. This is not a trustworthy method as the results are liable to be easily manipulated.

(2) **Relatives of Aggregates Method**:—As in the previous method totals of the prices of the various commodities are taken but they are turned in to percentages by taking any particular year to serve as the base. Assuming the price totals of the different years in the previous example to be Rs. 91–6–0 in 1930

Rs. 95–2–0 in 1932 and Rs. 93–7–0 in 1934 the results will be expressed as under by taking 1926 as the base year:—

Year	Total Rs.	Percentage
1926	83-10-0	100.0
1928	79-13-0	95.4
1930	91- 6-0	109.3
1932	95- 2-0	113.8
1934	93- 7-0	111.7

This method is a little better than the previous one but even then it is far from satisfactory.

(3) **Average of Relatives Method:**—In this method percentages are taken of each item based on the questions of the base year or period and not on totals as in the second method. The average is then taken of the total and it will give the Index Number of Prices.

Commodity	Unit	1925	1937	Percentages
		Rs.	Rs.	
1. Bajra	Per maund.	1–8–0	1–15–0	129.2
2. Cloth	Per yard	0–4–0	0–3–6	87.5
3. Salt	Per 5 seers	0–2–0	0–2–3	112.5
4. Oil	Per seer	0–2–6	0–2–0	80.0
5. Fuel	Per maund	0–11–0	0–12–0	109.1
6. Spices	Per seer	0–1–0	0–0–9	75.0
7. Tobacco	Per bundle of Bidis	0–0–9	0–1–0	133.3
8. Rent	Per room	3–0–0	4–0–0	133.3
Total				859.9

The Index number of Prices for the year 1937 based on the prices of 1925 is 107.48 or 107.5 and is obtained by dividing the total of 859.9 by 8 i. e. the number of items taken. If this example is worked out by the method of aggregates the total of prices for the year 1925 comes to Rs. 5-13-3 and that of 1937 to Rs. 7-4-6. By adopting the preceding method the percentage worked for Rs. 7-4-6 on the base of Rs. 5-13-3 will come to 124.1 while our result has been 107.5.

(4) **Chain Base Method:**—The base in this method is not fixed but the preceding year is taken to serve as base.

Chain Relative Index Numbers (1935=100)

Year.	Index Number of Commodities						Average	Chain index number
(1)	A (2)	B (3)	C (4)	D (5)	E (6)	Total (7)	(8)	(9)
1935	100	100	100	100	100	500	100	
1936	93	112	87	94	104	490	98.0	
	93	112	87	94	104	490	98.0	98.0
1937	97	107	94	90	126	514	102.8	
	104.3	95.6	108.0	95.8	121.1	524.8	104.9	102.8
1938	104	114	98	93	119	528	105.6	
	107.2	106.5	104.2	103.3	94.5	515.7	103.1	105.9
1939	112	118	102	99	108	539	107.8	
	107.6	103.5	104.0	106.4	90.9	512.4	102.4	108.4
1940	105	112	107	105	103	532	106.4	
	93.8	95.0	104.9	106.0	95.4	495.1	99.0	107.3

In the above example the Index Numbers of prices of the various groups of articles, A, B, C, D and E are given for different years as worked out by taking 1935 as the base year. In column no. 8 are given their average i. e. the Index Numbers of prices constructed by the fixed based method. In column numbers 2,3,4,5 and 6 are given (between the different years) percentages worked out on the basis of the change of the Index Numbers from the preceding to the succeeding years,, For instance, in 1937 the Index Number for A is 97 while in 1936 it was 93 giving therefore ehe percentage of 104.3. Such percentage figures have been obtained for B, C, D and E which work out to 95.6, 108.0, 95.8 and 121.1. All these are the percentages calculated on the bases of 1936. The total of all these percentages as entered in column no. 7 is 524.8 and gives and average 104.9, when divided by 5 i. e. the number of items taken. The average percentage thus obtained is 104.9 and so with this the chain base Index Number of the previous year i. e. 98.0 comes to 102.8. In the same way the chain base Index Number can be calculated for the year 1938, As before the percentage for the various commodity groups is obtained on the basis of the change from 97 107, 94, 90, and 126 to 104, 114, 98, 93 and 119 respectively. The percentage figures thus arrived at are 107.2, 106.5. 104.2, 103 3 and 94.5 and so the average, 103.1 has been given in column 8. With this percentage of 103.1 the chain base Index Number for the year 1938 on the base of the preceding year i. e. 102.8 comes to 105.9 (100 : 102.8 : : 103.1).

The special advantage of this Chain Base Method is that it is possible to make a direct comparison between each year and the succeeding or the previous one which is more useful than the indirect comparison with some remote year taken as base. Again, it is easy to drop old items or to add new ones without much difficulty. This method, however, involves more labour and does not afford facilities for making comparisons.

Uses of Index Numbers:—In the beginning, Index Numbers were mainly used for measuring changes in the price level. Prices of all articles do not move in the same direction and even, if by chance they do, they rarely move with the same velocity. It, therefore, becomes necessary to ascertain the general level and this is achieved by computing Index Numbers. They are found to be particularly useful in estimating the cost of living. Cost of living Index Numbers show us whether real wages are rising or falling with the money wages or even if they rise whether the rise is proportionate or not. They are thus of great use in fixing allowances and bonus to the workers and thereby help us in settling the vexed question of wages, the most important economic problem of the day. Even the Pay Commission of the Government of India has based its recommendations on the Index Numbers.

Industrial activity can be studied with the help of indices of industrial activity and they will also serve as a guide to the Tariff Board to see whether protection given to a particular industry has been fully justified or not by its results. To those in the Stock Market Investment Index Numbers will help considerably in dealing with stocks and securities. Similarly, various other index numbers computed for unemployment, deferred payment and volume of trade will be of much use in tackling different economic problems. They are, therefore, rightly called 'Economic Barometers'.

It must however, be noted that very often Index Numbers computed for one purpose are used for a different one without proper precaution and this gives rise to some erroneous interpretation of facts. For instance, Index Number of prices worked out for studying conditions of labouring classes cannot, without adjustments, be used in the case of upper classes Though an Index Number will give a better idea of the situation than figures of absolute rise or fall it must be noted that it can be trustworthy in its tendencies wtthout being reliable to the last digit. It is no doubt true that they provide a means whereby the trend of events relating to certain data can be studied with great ease particularly with their graphical representation and they are likely to be more reliable when they are based on a fairly large number of items.

Various Index Numbers:—The computation of Index : Numbers, as already explained, involves various problems which are tackled differently by different persons. With a view to have a comprehensive idea of the various problems, working of some of the important indices in this as well as in other countries has been briefly described in the following pages.

Index of Business Activity in India·—This Index Number of business'activity is being computed and published every month from March 1938 by the well-known weekly paper 'Capital' of Calcutta. The technique of its construction was explained in detail in the issue of 'Capital' dated 17th March, 1938. The year 1935 has been taken as the base period and series of seasonal fluctuations were corrected by applying twelve months moving average over a period 5½ years i. e. from April 1932 to

September 1937 and the geometric mean is used and not the arithmetic mean, though it is usually made use of in the construction of indices. The weightage assigned to the different items is as under :—

I.	Industrial Production.		
	(a) Cotton manufactures.	9	
	(b) Jute manufactures-	6	
	(c) Steel ingots.	5	
	(d) Pig iron.	8	
	(e) Paper.	5	
	(f) Cement.	3	
			36
II.	Mineral Production. Monthly raisings of coal only		7
III.	Index of Internal Trade.		24
IV.	Financial statistics— Monthly cheque clearances.		20
V.	Indices of values of Exports and Imports—		
	Exports.	4	
	Imports	3	7
VI.	Index of shipping activity.		6
			100

The figures of industrial production are taken from 'Monthly Statistics of Production of certain selected industries of India' and those of mineral production from the 'Trade Journal'. Index of Internal trade is based on monthly figures obtained from 'Accounts relating to the Inland Rail and River borne Trade of India.' Cheque clearance figures are taken from the Statistical summary of the Reserve Bank of India. Indices of the value of exports are taken from 'Accounts relating to the Coasting Trade'. Figures for shipping activity are based on the 'Monthly Accounts relating to the Coasting Trade and Navigation of British India' and on 'Monthly survey of Business Conditions in India.'

Working Class Cost of Living Index Numbers for Bombay City :—An Index Number of Cost of Living is the ratio that the current cost of a fixed collection of goods and services bears to the cost of the same collection of goods and services during the base period. In India official index numbers of working class cost of living are compiled and published from month to month for various provincial centres. As all of them are computed on identical lines with some modifications the working of the Index Number for the Bombay city has been described at length to serve as an illustration.

The Working Class Cost of Living Index Number for Bombay City was first published in 1931 and is being monthly compiled and published by the Bombay Labour Office in the monthly *Labour Gazette* issued by the Directorate of Labour Information, Government of Bombay. To start with this Index Number was based on Aggregate Consumption Method but this has been given up in preference to the Family Budget Method. The results of the first enquiry into working class family budgets in Bombay City between May 1921 and April 1922 has been superseded by those of the second enquiry between May 1932 and June 1933. This Index Number which is worked by taking the arithmetic average has for its base the average of the period of twelve months from July 1933 to June 1934.

The items taken for the revised Index Number like the cost of living Index Number of the British Ministry of Labour are classed into five groups;—(1) Food, (2) Fuel and Lighting, (3) Clothing, (4) House Rent and (5) Miscellaneous and weightage is given in the following manner to the items of these groups.

(1) Food—Rice 22 ; Patni 6 ; Wheat 3 ; Jowar 1 ; Bajri 4 ; Turdal 4 ; Gram 1 ; Raw-Sugar (gur) 1 ; Sugar (refined) 5 ; Tea 2 ; Fish (dry Bumlows) 3 ; Fish (fresh Bhing or Palah) 1 ; (fresh Prawns) 2; (Fish fresh Bumlows) 2; Mutton 5; Milk 7; Ghee 2;

Salt 1; Chillies (dry) 3; Tamarind (old) 2; Turmeric 2; Potatoes 1; Onions 1; Brinjals 5; Pumpkin (white) 5; Coconut oil 2; Sweet oil 2; Tea (ready made)5.

(2) Fuel and Lighting—Charcoal 30; Firewood 52; Kerosene oil 16; Matches 2.

(3) Clothing—Dhotis 15; Coating 12; Shirting 23; Cloth for trousers 4; Sarees 36; *Khans* 10.

(4) House Rent—100.

(5) Miscellaneous—Barber 13; Soap (washing) 9; Medicine 3; *Supari* 25; Bidis 22; Travelling to and from native place 27; Newspaper 1.

It will be seen from the above that the total weightage for each group is 100 but the number of items in each case varies considerably. Even in the case of Index Numbers of three main centres in this Presidency there is a great variation in the number of items in the groups as will be seen from the following:—

Items in different groups

Groups	Bombay city	Ahmedabad	Sholapur
Food	28	16	17
Fuel and Lighting	4	4	3
Clothing	6	7	6
House Rent	1	1	1
Miscellaneous	7	2	5
Total	46	30	32

Besides giving weightage to the items in the various groups it is also applied to the groups and that too differently for different centres.

Weight Proportional to Total Expenditure.

Groups	Bombay	Ahmedabad	Sholapur
Food	47	58	49
Fuel and Lighting.	7	7	10
Clothing	8	10	12
House rent.	13	12	6
Miscellaneous.	14	4	6
	89	91	83

It will be seen from the above that not only the number of weekly items in each group is different for these three centres but weightage applied to index numbers of the groups to compute the index numbers for the various cities varies considerably. This is due to the findings of the inquiries instituted at these places. Both at Ahmedabad and in Bombay 3 percent random samples of tenements were collected without any income limit. In Ahmedabad 872 budgets were accepted in 1926 while for Bombay Enquiry in 1932–33 the number was 1469. In Sholapur too 902 budget were accepted in 1925. Before 1933 the weightage applied to the groups used to be 100 but as a result of the budget enquiry the weightage in the case of Miscellaneous (Bombay Centre) brought down to 14 from 25 and so the total of the weightage has come down to 89 from 100.

In Bombay the prices for almost all articles except clothing, fish, brinjals and pumpkins are collected weekly by officers of the Labour office from two shops in each of the twelve different industrial areas. The prices of cloth are taken from the retail shops of four mills. Since June 1943 the calculations in Bombay city have been based on the prices of the different cereals sold in government grain shops and authorised ration shops in working class localities. This is a temporary measure and as soon as the control is removed market quotations will be obtained for compiling the Index Numbers.

Economic Adviser's Index Numbers:—The Index number of Wholesale Prices for certain articles in India, better known as Economic Adviser's Index Number, appeared for the first time in June, 1942. It is compiled every week and generally monthly averages are also given. The week ending 19th August, 1939 is taken for its base and the geometric mean is used in its computation. This index is based on the prices of 23 articles classified into four main groups: (1) Food and tobacco, (2) Other Agricultural commodities, (3) Raw materials and (4) Manufactured Articles. In food and tobacco group are included eight items. Cotton, jute and linseed are the three items in the second group while in the third and fourth groups, items taken are seven and five respectively. A sensitive index, called the 'Primary commodities Index', is also compiled by taking the prices of the first three groups, in all eighteen in number. The other index computed is 'Chief articles of export' and the prices taken for this are all–India wholesale prices.

All-India Index Numbers of Wholesale Prices of Food Articles (new series) are also compiled by the Economic Adviser. The weightage for this is Rice 36, Wheat 12, Jowar 8, Bajra 3, Gram 5, Dal 3, Tea 7, Coffee 1, Sugar 13, Gur 9, and salt 3, in all 100. Spices and condiments are not included.

The Government of India have formulated a scheme for the preparation and maintenance of cost of Living Index Numbers on a uniform basis because of the unsatisfactory character of the retail price Index Numbers as pointed out by the Rau Court of Inquiry appointed to investigate into the dispute regarding dearness allowance on the G. I. P. Railways. The committee remarked, "none of the cost of living index numbers at present available are entirely satisfactory." The scheme was discussed at the Third Conference of Labour Ministers held in Delhi in January 1942 and the important conclusion reached is that there should be uniformity of technique in the computation of cost of living Index Numbers in the various provinces.

The Government have also decided in consultation with the Provincial Governments to proceed concurrently with three other schemes : (a) The Main Cost of living Index Number scheme (b) Retail Price Index Nnmber scheme for urban centres and (c) Retail Price Index Number scheme for rural centres.

Cost of Living Index Number (British Ministry of Labour)-This Index Number is compiled by the British Ministry Labour with a view to measure the average increase in the cost of maintaining the pre-war (1914) standard of living of the working classes. It was on the basis of the Index Number on which claims for increase in wages were made by Trade Unions during the War. Many disputes relating to wages are settled by adjusting the basic rates of wages by a sliding scale depending upon the Index Number.

The items taken are classed into five main groups, viz: (1) Food, (2) Rent, (3) Clothing, (4) Fuel and Light and (5) Sundries. The information about food items is obtained by addressing enquiries to over 5000 retailers enjoying a working class trade in over 500 towns and villages. The weights applied are on the basis of the average expenditure of 1944 urban families of the working class as ascertained in 1904 by the Board of Trade. Prices in July, 1914 are taken as the base for this index. The cost of living index number on the 1914 base was calculated for the last time in respect of June, 1947. As from that date a new index of retail prices has been calculated with prices at June 17 taken as = 100 as an interim measure pending further study and examination. The items of food taken cover over three-fourths of the total family expenditure on food and the items, not included because of wide variations in quality and seasonal supplies, are fruits and vegetables other than potatoes. The items of food and weights applied are as under:—

Beef	48	Fish	9	Tea	22	Butter	41	Eggs	19
Mutton	24	Flour	20	Sugar	19	Cheese	10	Potatoes	11
Bacon	19	Bread	50	Milk	25	Margarine	10	Total	327

Rent figures are obtained from Local Authorities as well as from Property Owners' Associations. Inquiries regarding clothing are addressed to outfitters. drapers, boot retailers, etc. over about 300 retailers in numbers and from over eighty towns. Coal merchants, gas suppliers and retailers are approached for figures for fuel and light and those for other items, which mainly include soap, soda, domestic ironmongery, brushware, pottery, tobacco. fares and newspapers, are got from retailers, transport undertakings and press reports.

In all cases retail and not wholesale prices are taken into consideration and such group is again given weightage on the basis of the Board of Trade survey made in 1904. They are meant to adjust the relative importance of such group and are as under :—

Food	7½	Fuel and Light	1
Rent	2	Other items	½
Clothing	1½	Total	12½

It will be seen from this system of weightage that food has been given the highest weightage, being 60 percent. The figure thus obtained is expressed in the form of an increase in the cost of living and to have an Index Number 100 will have to be added to it. The working class cost of living Index Numbers in India are worked out more or less on identical lines.

The Economic Index of Business Activity :—This Index Number is compiled every month by "The Economist' to measure the changes in the economic activity of United Kingdom and to show the trend of trade in quantitative units. It also gives an approximate idea of the variations in the *real* national income.

Its compitation was started as early as 1924 but it was revised in July, 1936 and was recalculated with 1935 as its base year. The composite Index Number is obtained by calculating the geometric mean of the various groups which are given below with the weightage applied in each case :—

		Weights.
1.	Employment	10
2,	Consumption of coal	4
3.	Consumption of electricity	2

		weights
4.	Railway merchandise tariff.	4
5.	Commercial vehicles in use.	2
6.	Postal traffic receipts.	3
7.	Building activity.	2
8.	Consumption of iron and steel.	2
9.	Consumption of cotton.	1
10.	Imports of raw materials.	2
11.	Exports of British manufactures.	3
12.	Movements of Shipping.	2
13.	Metropolitan, Country and Provincial Bank clearings.	4
14.	Town Bank clearings.	1

All the series except three items relating to employment, motor vehicles and building activity are on a daily average basis minimising thereby the effect of months with different number of days.

Investment Index Number :— This Index Number, compiled by the Bankers' Magazine in Great Britain for the benefit of those interested in the Stock and Share Market, gives the estimated level of the prices of stock and shares. The quotation of 365 stocks and shares grouped in two classes on 31st December, 1921 have been taken to form the base=100. The two classes consist of 87 securities of fixed interest stocks and 278 securities yielding variable dividends. Adjustments are made on fresh issues and on liquidation of a company. Average market values can be compared with the base to obtain the rise or fall of investments.

The Investors' Chronicle also compiles its Investment Index Number by taking 31st December, 1923 as the base=100. For this, 176 quotations of market values of securities are taken and the grouping is differently made. In the fixed interest group are included 34 securities of which 19 are gilt-edged and 15 relating to business. In the other group the number of variable dividend securities is 142 of which 16 are of first-class securities, 70 of general business securities and 56 of speculative type.

Separate Index Numbers are also compiled for each group. In this the base year taken is a little far from the period of depression following the post–war trade boom as is the case of the Bankers' Magazine Index Number.

Index of Cost of Living (U. S. Bureau of Labour) :— The united States Bureau of Labour compiles the cost of Living Index Number and it is published in the monthly 'Labour Review'. For this the year 1913 has been adopted as the base and the compilation of this Index Number was begun in the year 1918 by computing the Index Numbers from December, 1914. Price quotations are obtained from markets and also from Bureau's field agents for commodities used by working class families. The items are divided into six groups, (1) Food, (2) Clothing, (3) Rent, (4) Fuel and light, (5) Furniture and furnishings and (6) Miscellaneous items and not into five as is done in the case of the Index Number of the British Ministry of Labour. The weightage is based on the study of 12,000 family budgets in U. S. A. and is also done for the items of the group. These indices are regularly computed for the different cities as well as for the country as a whole.

Dun's Index Number :— This Index Number is regularly published by a private organisation about the middle of the month in Dun's Review of United States. This was first published in 1901 and is a weighted aggregate of actual wholesale prices at the beginning of the month. The weightage is on the average annual per capita consumption and food items are given nearly fifty per cent importance. Over two hundred commodities are taken and they are grouped into seven classes, their wholesale prices being obtained from the principal markets of U. S. A.

Exercises.

1. What is an Index Number? State reasons for using Index Number in the beginning primarily in the case of prices.
2. State briefly the problems that arise in constructing an Index Number of Prices.

3. Give the different methods adopted for constructing an Index Number of Prices and describe fully the method you think most suitable.

4. Write short notes on the choice of items for constructing an Index Number of Prices and also on the methods of giving them weightage.

5. Write a critical note on the construction of an Index Number of cost of living. (1941–Mar.)

6. What is meant by weighting in statistics? Illustrate by reference to averages and Index Numbers. (1944–Mar.)

7. In what sense is it possible to measure a general level of prices by means of an Index Number?

8. Mention the chief well-known Index Numbers in England and in India Explain the details of the construction of any one of them. (1932–Oct.)

9. What is the objective of an Index Number? State briefly the relevant conditions for its construction illustrating your answer by reference to the Index Number of prices published by the Government of India. (1936–Mar.)

10. Define the term 'Cost of Living' and point out difficulties encountered in measuring the change in the cost of living over a period of time. (1942–Mar.)

11. 'The significance or validity of any Index Number depends upon the dispersion of the price relatives upon which it is based.' Comment. (1943–Mar.)

12. Write short notes on chain base method, inverse prices, relative of aggregates method, and economic barometer.

13. Mention the chief sources of error in the Cost of Living Index Number and state how they can be avoided.

14. Distinguish between the fixed Base and Chain Base methods of constructing an Index Number. Discuss the relative merits of each method.

15. Explain fully the procedure adopted in computing the Cost of Living Index Number of the British Ministry of Labour.

16. What is the purpose of an Investment Index Number? Describe the working of any Investment Index Number you know of.

17. From the following table giving average prices of twelve important commodities at Agra, prepare Index Numbers of prices for the years 1926 and 1928 based on the average prices of the year 1925.

		Average Prices in 1925 Rs.	Prices in 1926 Rs.	Prices in 1928 Rs.
Rice	Per Maund	5–14–0	6—2–0	6—8–0
Wheat	,,	5—4–0	5—2–0	5—6–0
Ghee	,,	65—0–0	63—0–0	61—8–0
Milk	,,	5—0–0	4–12–0	4—8–0
Fuel	Per Candy	27—8–0	26—4–0	25—9–0
Salt	Per Maund	4—3–0	3–15–0	3—9–0
Potatoes	,,	1–14–0	2—3–0	2—1–0
Sugar	,,	14—4–0	13—8–0	12–12–0
Cloth	Per Yard	0—7–0	0—8–0	0—9–0
Oil	Per Maund	5—0–0	4—8–0	4—3–0
Bajra	,,	3–12–0	3–15–0	3—4–0
Pulses	,,	5—8–0	6—3–0	6—7–0

18. Describe the chain Base method in the construction of Index Numbers from the following table and calculate such Index Numbers.

Year	Articles (Index Nos.)			
	A	B	C	D
1925	94	87	81	114
1926	98	83	84	119
1927	107	89	82	122
1928	94	82	97	107
1929	96	84	96	105

19. The following are the average prices of some commodities in U. P. during 1925 and 1928. Calculate the general Index Number of prices based on the prices in 1925.

Commodity	Unit	Average Prices	
		in 1925	1928
Wheat	Per maund	5—4—0	6—8—0
Cloth	Per yard	0—6—6	0—9—0
Salt	Per maund	4—0—0	3—9—0
Sugar	„	13—5—0	12-12—0
Ghee	„	64—4—0	61—8—0
Milk	„	5—2—0	4—8—0
Rice	„	6—4—0	6—8—0
Oil	„	5—3—0	4—3—0
Pulses	„	5-12—0	6—7—0
Fuel	Per candy	26—8—0	25—9—0
Potatoes	Per maund	1-15—0	2—1—0

20. The group Index Numbers of different items are given below and you are asked to construct an Index Number of prices for June 1937 by assigning the following weights.

Food	28	Fuel and lighting	4
Clothing	6	Rent	1
Miscellaneous	7		

Group Index Numbers of June 1937.

Food	112	Fuel and lighting	169
Clothing	84	House Rent	100
Miscellaneous	97		

21. From the following information prepare the Index Number for the Miscellaneous Group taken for the construction of the New Bombay working class cost of living Index Number.

Miscellaneous Group

No.	Articles.	Unit of quantity.	Weight assigned.	Average price in the year ended on 30-6-34 Rs.	Price in June 1937. Rs.
1	Barber	Per shave	13	0-1—4	0-1—0
2	Soap	Per bar	9	0-4—5	0-4—0
3	Medicine	Per bottle of mixture	3	0-8—0	0-8—0
4	Supari	Per lb.	25	0-5—0	0-5-11
5	Bidis	Per bundle of 25	22	0-1—0	0-0-11
6	Travelling Expenses	...	27	0-4-11	0-4—8
7	Newspaper per copy		1	0-0—9	0-0—6

CHAPTER XV

INTERPRETATION OF STATISTICAL DATA

After the data have been collected and presented by textual, tabular or graphic methods the next step is the analysis and interpretation of the statistics compiled. Statistics, as they exist or after they are tabulated and summarized, may give an idea of what has happened but a logical interpretation by an expert would help us in understanding the underlying process at work as well as their implications. If statistical records of rainfall are presented we would get an idea of the rainfall in the past but a clever statistician, after statistical operations, would be able to discover years of plenty and draught as well as their likely periodicity. He would also be in a position to forecast, which in the absence of *abnormal* circumstances, is likely to serve as a guide for future years. A Statistician is, therefore, expected not only to assemble and analyse the data but to interpret the result of his findings as that is very often the main purpose underlying the collection of data. This work of drawing inferences from statistical data or their interpretation is of a technical nature and, if left to inexperts, is likely to lead to misuse of statistics even without any motive. Any person without adequate grounding in scientific method as well as statistical operations and their implications is apt to draw wrong conclusions. Simply because the coefficient of correlation shows correlation between prices and number of deaths in a locality it is not to be concluded that there exists correlation between the two. Common sense as well as experience are also necessary for such a work. In addition to this it is necessary that the whole work connected with the particular problem should be entrusted to the same person and if this is not practicable due to heavy work at least it should be under his supervision and guidance. In such a case, the statistician will be fully conversant with the problem and its implications and the inferences drawn under such circumstances will be more precise as well as reliable. The Census Commissioner, who planned the census procedure and supervised it, will be in a better position to interpret the census figures than one who takes them from the census reports.

In conducting scientific experiments with a view to interpret their results and to draw out inferences therefrom it is possible to allow at one time only one factor to vary by keeping others under control by scientific methods. In the case of social sciences the application of the Law of Single Variable is out of question as it is not possible to check multiple variation though in some cases their free play can be minimised. This difficulty is to some extent got over by the use of scientific methods which according to Yule are "especially adapted to the elucidation of quantitative data affected by multiplicity of causes." This is in fact the fundamental distinction between the experimental and statistical methods. The former are manipulative while the latter are descriptive.

Pre-requisites of Interpretation:—As already remarked the work of interpretation is very difficult and needs great care and minute attention to details. With a view to be able to draw reasonable and correct inferences it is imperative that as preliminaries to interpretation the following considerations should be borne in mind.

(1) **Adequacy:**—Our inferences are based on the data supplied and it is, therefore, necessary that they are adequate to base our judgments upon. Not only that the data should be properly selected but the number of observations taken must be large enough to present a true and correct picture of the universe or population from which they have been drawn. It is not unusual to find data collected by deliberate sampling and not by random sampling. In such a case it is but natural that if the investigators are biassed and they are consciously or even otherwise likely to take such observations as are in the conformity with their views. In the case of random sampling this danger does not exist as the persons concerned have no choice of selection. It is also necessary that an adequate number of cases must be taken as by taking a small number the average will not conform to the general universe or population as the influence of the Laws of Statistical Regularity or Inertia of Large Numbers is visible only when the number is sufficiently large.

(2) **Stability:**— At the time of accepting the data for interpretation it must be seen that there is likely to be stability of the data. The test for this is that the results must repeat in similar experiments. If this is not achieved there is no stability and inferences drawn from such material will serve no useful purpose.

(3) **Homogeneity:**—The data must also be homogeneous and comparable. When cases are being compared for drawing inferences it is futile to do so without homogeneity. When wages in two concerns are being considered it is necessary to see that other things such as nature of work, leisure and other facilities granted are identical. If bad debts are being reviewed with a view to draw some conclusions it would be a mistake to compare the percentage of bad debts in two concerns when in one they are calculated on sales consisting of cash as well as credit and in the other on credit sales only. In short, there must be logical consistency.

(4) **Relevancy:**—The data taken for interpretation must be relevant. There must exist some significant connection between the purpose of enquiry and the results of the analysis and the drawing of inferences. In the absence of this condition the result arrived at will not be reliable.

(5) **Accuracy:**—The data must be free from errors of all types biassed as well as unbiassed and also those of manipulation e. g. unconscious errors in measuring, weighing, counting or approximation. Special care must be taken to see if there has been a possibility of cumulative errors in collection, as unlike compensating errors they go on accumulating errors.

(6) **Scientific Analysis:**—It is equally necessary that the data must have been scientifically worked out and analysed and as far as practicable disturbing factors reduced to minimum and in case this is not possible they are considered with all their complications.

By taking care of the above points the possibility of mistakes in drawing inferences will be considerably reduced and considerable loss of energy and time can be saved besides increasing the reliability of the inferences drawn.

Errors of Interpretation:—Errors in interpretation are due to various factors intentional or otherwise. There is an innate desire in man to interpret the data and in doing so very often he disregards his limitations. He tries to draw inferences without proper mental preparation as well as the requisite scientific training. It is not imperative that he must possess a degree or diploma in statistics but at least he must have the necessary grounding in the subject coupled with sufficient experience of handling such problems. In the absence of this it is very likely that he will jump to conclusions not warranted by the data.

The other factor against which we must particularly guard is the deliberate misuse or misinterpretation of statistical facts by interested parties such as election against, advertisers and propagandists. These persons with a view to serve their purpose use inaccurate or incomplete data, misleading methods of presentation, false generalizations and faulty use of statistical methods. It is argued that wages were lowered by twenty percent and were later on raised by twenty percent so the labourers should have no cause of complaint. But on careful consideration it will be noticed that it is not so. A labourer who was getting originally Rs. 50 per month because of this reduction of 20% began to get Rs. 40. Later on when wages were raised by 20% his monthly income came to Rs. 48 and not Rs. 50 which was his original monthly income. The propagandist in favour of employers tried to make out a case by saying that the percentages of rise and fall are the same and so there is no possibility of any reduction in wages. Similarly, interested parties having an even diametrically opposite interests try to prove successfully from the same data that their point of view is correct. One interested in showing that the rainfall of a particular year is satisfactory will compare it with a year of draught in the past while the one trying to prove it to be far less will choose for comparison a year of having a very heavy rainfall. Thus by shifting the base of comparison the conclusions drawn appear to be absolutely different and opposite in character. False generalization is very commonly resorted to for misinterpreting things. Simply because of the increase of imports it would not be safe to conclude that the country has grown rich as if otherwise it would not have been in a position to have this rise in imports. It may be likely that its internal production has gone down and so it was required to import to satisfy its legitimate wants. If in a big city the death rate is found to be lower it may not be quite correct to conclude only because of this that it is healthier than a village. It often happens that only able bodied persons who are capable of earning wages stay in the city and the old and sick persons go back to their villages, as it found difficult for them to maintain themselves in big cities without work due to higher cost of living.

Errors in interpretation also arise because of the incorrect or faulty data. This generally happens in the case of based investigators. To establish their point of view they resort to deliberate sampling which as already shown is a faulty method of collecting data. They also do not gather sufficient data but are deliberately content with a small number of observations favouring in most cases their point of view. This does not allow a free play of the law of statistical regularity and thereby fails to present a true and correct picture of the universe.

Statistical methods, if not used in the proper way, are likely to lead to wrong interpretation of the data. While constructing Index Number of prices different results are obtained by using different weights or by taking quotations of prices of the same article but of different qualities. The base used can be changed and it does materially influence the Index Number obtained. Similar mistake is possible by wrong interpretation of the coefficient of correlation or by not recognizing 'nonsense correlations'. Simply because there is high degree of correlation between the figures of imports and those of birth rates it cannot be concluded that there exists a casual relationship between the two.

General Directions:—The work of interpretation or of drawing inferences from statistical data is of a complicated nature and as it is not possible to lay down a set of rules or formulas for this purpose some general directions have been given to serve as a guide to a beginner.

To start with the available data should be subjected to the various statistical operations. By ascertaining the various forms of averages such as mean, median and mode a concise picture of the groups can be obtained and this will help in making comparisons with other groups of a similar type as well as for judging the relative positions of the various items in the group. In the case of historical or time series it will pay to first analyse carefully the influence of the various types of movements like the long-time trend, seasonal variations, cyclical movements and irregular fluctuations by resorting to one of the methods already explained. Even simple moving average will be found very useful. By measuring dispersion and skewness we shall be in a position to see how far the various items differ from one another in size and if there is a symmetry in the group or not.

Graphs, if drawn, will be very useful and thereby we will be able to easily notice certain tendencies in the data as it is difficult to do so by merely scanning the tables. To study the existence or otherwise of unequal distribution Galton's graph can be made use of. An idea of the future development can be obtained by extrapolation in case the data is of a normal type. Again, if we have to study two or more groups or series it will be worth while to calculate the coefficient of correlation or to represent it graphically to see if there exists correlation between the two groups or series. It is equally necessary to see if there is a likelihood of a lag in the given data.

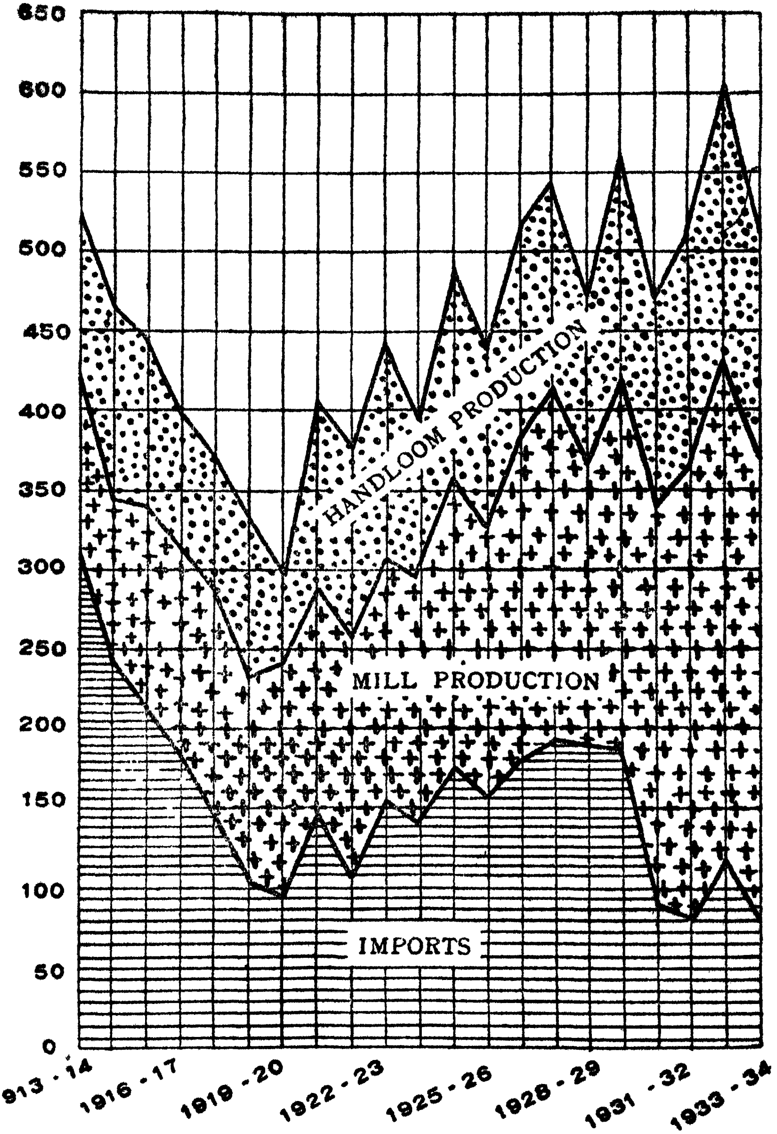
It will be a serious mistake if the data is interpreted without studying the back ground as well as the general implications. It is very likely that circumstances might have changed materially in the period under consideration. In interpreting the import and export figures of a country for a period it is absolutely necessary to see if there have been material changes during the period. The separation of Burma and Pakistan from India is bound to influence the figures due to the reduction in the area and population of the country. Again, if there are years of war, pestilence or famine, trading in those years is sure to be affected and if these circumstances are not borne in mind the interpretation will be incorrect. Lastly, it is of vital importance to note that statistical laws are true in the case of averages and not necessarily in the case of every individual item. The interpretation, therefore, should not be dogmatic in character and should not be made without bearing in mind the various pitfalls already explained in the foregoing pages.

Position of Cloth in India (in crores of yards).

Years	Net imports	Net available mill production	Handloom production	Net available for consumption
1913–14	313	108	107	528
1914–15	241	107	118	466
1915–16	211	133	105	449
1916–17	183	132	82	397
1917–18	147	142	81	370
1918–19	101	130	105	336
1919–20	99	144	56	299
1920–21	145	143	115	403
1921–22	102	157	119	378
1922–23	152	156	134	442
1923–24	142	154	101	397
1924–25	177	179	126	482
1925–26	153	174	116	443
1926–27	176	206	133	515
1927–28	194	219	131	544
1928–29	191	174	108	473
1929–30	190	229	140	559
1930–31	87	246	139	472
1931–32	76	288	150	514
1932–33	120	311	170	601
1933–34	77	289	144	510

If we were to interpret the above statistical table showing the position of cloth in India for a period of 21 years it would facilitate our work considerably if the data is subjected to statistical operations including its graphical representation preferably by drawing band or belt curves as shown in the given figure.

POSITION OF CLOTH IN INDIA



This table, comprehensive as it is, gives in crores of yards imports of cloth from outside and the net available mill and handloom production as well as the total of the net available cloth for consumption for the country. This data covers a period of 21 years i. e. from 1913–14 to 1933–34 in which is included the abnormal period of the Great War from 1914–18. During the war and few years, thereafter, the economy of a country is far from normal. The foreign trade is particularly affected, its extent depending upon the position of the countries at war. This can be clearly seen by scanning the import figures of cloth. From 1914–15 there appears to be a regular decline up to the year 1919–20. As the supply of cloth from outside during this period went down the mill production improved to meet the demand in the country. The handloom production showed a rise just at the beginning of the war period but mainly because of the competition by the mills the handloom production got a setback.

It was only in 1919–20 that the net available amount for consumption went as low as 299 crores of yards as against 456, being the average for this period of 21 years. This is due to all round decrease just at the cessation of the war. Due mostly to the non-cooperation movement the imports in 1921–22 went down and the mill as well as handloom production advanced to meet the demand. The slump period after 1929 coupled with the nation-wide *swadeshi* movement affected the imports which touched the lowest point of 76 crores in 1931–32 as against the average of 156 during the whole period under review. Due to the national awakening the production at home got a fillip and this is visible from the rise in mill and handloom production till it was arrested again by a slump after 1932.

From the graph it can be clearly seen how during the war period the imports have decreased and though they made a little headway they again went down in later years due to slump as well as national movement. Except for the few years of war the net available amount for consumption has been showing steady rise but for a few years of slight reverses. The handloom production during the war years was fairly steady and thereafter has improved to a great extent. In the case of imports we find the coefficient of dispersion very high as compared to those of the handloom production and of the net available amount for consumption. By calculating the coefficient of correlation or even by scanning very carefully the movements of the various curves, positive correlation of a fairly high degree is seen between the mill and handloom production because of their rising production due to fall in imports and the national movement. There appears to be inverse correlation but of partial nature between the imports and mill production because of their movements in opposite directions—mill productions are rising while imports are falling.

It will be seen from the above example how we are enabled to interpret the statistical data better with the help of statistical operations. It is, however, needless to emphasise that a fair amount of knowledge of the background of the data and absence of bias are absolutely essential.

Exercises.

1. What is meant by interpretation of statistical data and what is its importance in the study of statistics?

2. Describe fully the errors that are likely to affect the interpretation of data and state how they can be taken care of.

3. Mention the important pre-requisites of interpretation and explain any three of them by giving suitable examples.

4. What points should be borne in mind in interpreting the statistical data given to you?

5. Write a brief but reasoned note based on the table giving age-distribution per thousand of the population in 1911.

Distribution of Population

Age.	Germany.	France.	Eng.	U.S.A.	Japan.	British India.
Under 10	234	171	209	222	244	276
10–20	203	166	190	198	198	192
20–30	164	158	173	187	154	178
30–40	139	148	152	146	138	142
40–50	105	127	115	106	101	99
50–60	76	104	80	72	77	61
60–70	51	77	51	43	57	36
70 & over	23	49	30	26	31	16

6. Comment upon the facts revealed by the following table giving percentages of exports of some of the principal crops of India to the total production.

	Pre-war average	War average	Post war average
Rice	9	5	5
Wheat	14	9	3
Tea	96	89	95
Cotton (raw)	56	51	61
Jute (raw)	51	31	48
Linseed	73	63	59
Rape and Mustard	23	8	19
Sesamum	25	8	6
Groundnut	35	12	19
Indigo	40	44	27

7. Interpret the facts contained in the following table:—

	Population in 1931 in millions	Mean density per sq. mile	No. of females per 1000 males
India	352.8	195	940
Ajmer Merwar	0.5	207	892
Assam	8.6	157	900
Andamans	0.02	9	495
Baluchistan	0.4	9	717
Bengal	50.1	646	924
Bihar and Orissa	37.6	454	1005
Bombay	21.9	177	901
C. P.	15.5	155	998
Delhi	0.6	1110	722
Madras	46.7	328	1025
N. W. F. P.	2.4	179	843
Punjab	23.5	238	831
U. P.	48.4	456	902

8. Comment upon the facts revealed by the following table.

BANKING

Country	Year	Banking Offices		Banking Capital & Reserve	Banking Deposits		Saving Bank Deposits	
		Total No.	Per million of population		Total	Per head of population	Total	Per head of population
				Rs. crores	Rs. crores	Rs.	Rs. crores	Rs.
India	1930	882	2.5	25.5	227.7	6.4	75.4	2.1
United Kingdom	1932	12557	273	316.5	3226.5	698	1232.9	266.8
U. S. A.	1931	22071	179	3540.6	13787.6	1123	6677.3	543.7
Canada	1931	4176	417	113.7	836	804	34.1	32.8
Germany	1928	2166	33	212.3	1084.8	167	1091.5	168.4
Japan	1929	6670	103	357.3	1407.5	215	581.6	88.9

9. Write a reasoned note based on the occupational statistics given in the following table:—

Occupational Statistics showing percentage of the total population working different occupations.

Country	Year	Total population in lakhs	Working population	Agri. Fishing	Industry Mining etc.	Trade and Transport	Public Instruction	Liberal Profession	Others
1	2	3	4	5	6	7	8	9	10
India	1931	3528	43.8*	29.4	4.4	2.9	0.3	0.7	6.1
Great Britain	1921	432	44.8	3.2	21.2	9.3	2 6	2.0	6.5
U. S. A.	1930	1228	39.8	8.8	12.6	9.7	0.9	2.8	5.0
Canada	1931	104	37.8	11.8	9.5	6.7	1.1	2.3	6.4
Germany	1925	632	50.7	15.4	20.9	8.3	1.0	2.5	2.6
France	1926	407	52.5	20.1	17.5	8.9	1.2	1.9	2.9
Japan	1930	645	45.3	22.8	8.9	9.1	3.2	0.1	1.2

*Two-thirds of the working population being agricultural has regular employment only for about half the year. The Column no. 4 gives the total working population of different countries while the succeeding columns show the distribution in various occupation.

10. Represent graphically the Index Numbers of wholesale prices of the following places and also write a reasoned note on the same :—

Year		Bombay*	Karachi*	Calcutta*
1923	...	181	148	172
1924	...	182	154	173
1925	...	163	151	159
1926	...	149	140	148
1927	...	147	137	148
1928	...	146	137	145
1929	...	145	133	141
1930	...	126	108	116
1931	...	109	95	96
1932	...	109	99	91
1933	...	98	97	87
1934	...	95	96	89
1935	...	99	99	91
1936	...	96	102	91
1937	...	106	108	102

*July 1914 = 100 (1939–Oct.)

11. Comment on the figures of the distribution of the Reserve Bank shares in the following table :—

	No. of shares in thousand		No. of share holders in thousands	
	1937	1938	1937	1938
Bombay	201	206	215	208
Calcutta	125	123	145	138
Delhi	94	93	157	149
Madras	60	59	91	87
Rangoon	19	18	18	16

12. **Draw up a brief but reasoned note on the prices quoted on the Bombay Stock Exchange from the following figures:—**

Name of the share or security	5-3-1937	5-2-1937	5-9-1936	5-3-1936
3½% Govt. paper	93—9	98—2	99-10	99—1
Bank of India	138—0	140-12	131—4	129—8
Tapti Valley Railway.	777—8	776—4	775—0	787—8
Bundi Cement ...	44—6	40-14	35—6	34—8
Bombay Dyeing...	970—0	895—0	654—8	667—8
Burma Trading Corporation.	695—0	651—4	489—6	368—2
New India	44-12	46—0	44—8	42—8
Tata (Ordinary)...	376—0	307—0	133-12	143—8
Tata (Deferred)...	1800—0	1402—8	717—8	516—4
Telephone Co. (Ordinary)	117—0	117—8	116—4	109—4

(1937-Mar.)

13. Comment on the results of workings of Class I railways in India after representing the matter graphically.

Year	Capital outlay in millions of £	Gross earnings in millions of £
1923-24	464	70
1924-25	473	74
1925-26	487	73
1926-27	505	72
1927-28	594	86
1928-29	599	86
1929-30	617	84
1930-31	627	77
1931-32	631	71
1932-33	638	70
1933-34	635	72

CHAPTER XVI

STATISTICS IN INDIA

For want of systematic historical records it is difficult to critically estimate the condition of statistics in Ancient India. However, interesting statistical information regarding army, production, earnings, etc. is got by referring to chapters 56–58 of Shanti Parva in Mahabharat, Jatak Kathas as well as Chanakya's Arthashastra. From 'Indika' and 'Anabasis' the two works of Arrian who came to India with Alexander the Great considerable statistical matter is available. The writings of Megasthenis, Greek Ambassador in India give very illuminating information on various topics such as census, army, births and deaths. The census procedure laid down is really admirable. There used to be an officer known as *Samahati* at the head like the present Census Commissioner. In the course of census operations information was collected regarding castes, occupation, houses, taxes paid in addition to counting the heads. Spies were engaged to report the infringement of rules and procedure. In the Mohamedan period too, there appears to have been very satisfactory Arrangement regarding the recording of statistics, Ain-i-Akbari by Abul Fazl contains a mine of information regarding the Mogul empire 'faithfully and minutely recorded in their smallest detail with such an array of facts, illustrative of its extent, resources, condition, population, industry and wealth as the abundant material from official sources could furnish'. H. S, Jarret, translator of Ain-i-Akabari has said of Abul Fazl, 'regarded as a statistician, no details from the revenues of a province to the cost of a pineapple, from the organisation of the army, and grades and duties of the nobility to the shape of a candlestick and the price of a curry-comb are beyond his microscopic and patient investigation'.

In the British period statistics of India, as remarked by the Bowley Robertson Committee 'originated as a bye-product of administrative activities and only in the case of population census and to some extent foreign trade there has been an organization.' It was in the year 1853 that the court of Directors in

England ordered the East India Company's office in India 'to collect statistics regarding native states, languages, opium, education, civil service, etc. In the early nineteenth century statistical data could be had only from the Annual Administration Reports published by the provinces but there was no uniformity in the method of collecting statistics. In 1862 forms were prescribed for compilation of financial and commercial statistics by the Statistics Committee. In 1868 the India Office published for the first time Statistical Abstract relating to British India for 1840–1865. Earlier in 1863 the Government of Bombay had asked for a special establishment for statistical work but this proposal was turned down.

The Moral and Material Progress Reports for India contain statistics regarding population, trade, production, prices as well as vital statistics but without the necessary explanatory notes. Gazetteers were published by the provinces in the nineteenth century and the Bombay Government issued matter covering over 18000 pages in thirty five books. Supplementary volumes were printed early in the twentieth century. The Statistical Atlas of the Bombay presidency was printed in 1888 but it was with a view to meet the demand of the Famine Relief Code to mark out districts of famine. By and by the collection of statistics was undertaken but there was no systematic and co-ordinated effort made in this direction.

The Indian Economic Enquiry Committee:–The Indian Economic Association in its session at Lahore in 1922 appointed a special committee to take up the question of approving the statistics. In February 1924 a resolution concerning this was passed by the Council of state and in Septomber by the Legislative Assembly. The Government of India then appointed the Economic Enquiry Committee in January, 1925 with Sir Mokshagundam Visvesvaraya as its Chairman and Prof. Burnett. Hurst as the Secretary. The terms of reference were 'to examine the material at present available for framing an estimate of the economic condition of the various classes of people of British India; to report on its adequacy; and to make recommendations as to the best manner in which it may be supplemented and as to the lines on which a general economic survey should be carried out with an estimate of the expenditure involved in giving effect to such recommendations.' This Committee's remarks on the then existing statistical material were that the general statistics other than those of production comprising of finance, population, trade, transport and communication, migration, and vital statistics were 'more or less complete.' Statistics of production including agriculture, pasture and dairy farming, forests fisheries, minerals, large-scale industries, cottage and small-scale industries were 'satisfactory in some respects but incomplete or totally wanting in others.' Statistics of income, wealth, cost of Living Index Numbers, indebtedness, wages and prices were not satisfactory and according to the committee no 'satisfactory efforts had been made in British India to collect the necessary material on a comprehensive scale. The committee recommended in particular (1) Quinquennial wages census for large industries, (2) preparation of estimates of total quantity and value of products of cottage industries and value of raw-materials consumed, (3) number of workmen employed, (4) establishment of a Central Statistical Bureau with a view to provide a common purpose and a central thinking office on the subject of statistics, (5) setting up of provincial statistical bureaus, and (6) legalising the statistical organisation thus ensuring and facilitating collection of statistics from individuals or firms. No serious effort was made to adopt these suggestions. In 1928 The Royal Agricultural Commission recommended the question of statistical organization but only a statistical section was opened for the Indian Council of Agricultural Research. In 1931, the Royal Commission on Labour also recommended the establishment of a statistical organization in each province.

The Bowley-Robertson Committee:—In August, 1933 the Government of India invited Dr. A. L. Bowley and Mr. D. H. Robertson to report, 'with a view to facilitating the further study of the economic problems of India, on the existing organization and range of statistical and other information bearing on the economic condition of India with special reference to the gaps which exist at present and the means of filling them and in particular to make recommendations as to the organization of a central statistical department which in addition to the collection of statistics can co-ordinate statistical enquiry and information for the whole of India, to consider the practicability and scope of a census of production, to consider the materials obtainable for measurement of national income and material wealth and to make recommendations as to the intention or continuation of series of Index Number of prices, wages, production, etc.' They came to India in December and after visiting important centres reported in March, 1934. This Bowley-Robertson Committee Report entitled 'A scheme for an Economic Census of India with special reference to a census of Production and Reorganization of Statistics' is a document of a very great importance, though probably because of its being compiled in a short period of even less than three months, it is rather brief and has failed to deal with statistics of metereology, literacy, education and public health except birth and death statistics. The following topics from this report are of very great importance and so they have been given at length in the following pages.

Organization of Statistics:—The Committee has rightly stated that 'the statistics of India have largely originated as a bye-product of administrative activities, such as the collection of land revenue, or from the need of information relating to emergency, such as famines. Only in the case of population census and to some extent of foreign trade there has been an organization whose primary duty is the collection of information. As a result the statistics are un-coordinated and issued in different forms by separate departments. Though in some branches careful work is being done and determined efforts made to improve the accuracy and scope of information, in others they are unnecessarily diffuse, gravely inexact, incomplete or misleading while in important fields, general information is almost completely absent. The only co-ordinated general publication is the statistical Abstract, which however, omits some important statistics which must be searched for in other documents. The situation cries out for overhaul under the control of a well-qualified statistician' !

To remedy this it was suggested that the title of the Director General of Commercial Intelligence and Statistics should be abolished as the work assigned to this post was too multifarious and a permanent economic staff consisting of four members attached to the Economic Committee of the Governor-General's Executive Council should be appointed. The four members to be taken should include one senior member who should work as the Secretary of the Economic Committee and should be in charge of preparing reports on urgent questions, the work of planning and general organization. Duties of the Director of Statistics, who should be a member of the Committee, should include (1) the conduct of population census, the office of which should not be temporary as at present, (2) the conduct of the Census of Production, (3) the co-ordination of Central Statistics and (4) the co-ordination of provincial statistics. The statistics branch of the Department of Commercial Intelligence and Statistics should be put in his charge with a view to facilitate his work. The Commercial Intelligence Branch, dealing mainly with the commercial enquiries, should be transferred to the Department of Commerce. The Director should be in contact with the other departments with a view to ensure uniformity and should be in charge of the publication of the Statistical Abstract. For each major province there should be a statistician to look to the compilation of statistics of the various departments and to conduct population census under the direction of the Director of Statistics. Till such offices are established in the various provinces the Director should arrange by contacting the provincial officers to promote uniformity in the provincial statistics with a view to facilitate their assembly into all-India totals. As the Census of Production is proposed to be quinquennial and as though the population census continues to be decennial, a supplementary census with an abbreviated schedule of enquiries mainly devoted to members, age, sex and occupation is proposed to be taken in the middle of the decennium there would be continuous work for the Director of Statistics. The remaining two members should be expert economists, working under conditions of utmost freedom which is compatible with membership of the civil service. As the field of economic work is so wide the senior and the other members should divide the territory for exploration amongst them according to their bent of mind and specialization.

Measurement of National Income:—The national income is the money measure of the aggregate of goods and services accruing to the inhabitants of a country during a year, including net increments to or excluding net decrements such as a loss of property by earthquake or disease to cattle from their individual or collective wealth. According to Dr. Bowley and Mr. Robertson the materials available for estimating the national income and wealth of India are very defective and the estimates so far attempted have become out of date thus requiring a fresh attempt *ab initio*.

Two methods of calculation have been given:—(1) the evaluation of the goods and services accruing (Census of Production method) and (2) a summation of individual incomes (Census of Incomes) method. The two methods do not furnish a check on one another over the whole field and for want of defective material in the country are not likely to yield very satisfactory results.

The first or Census of Production method involves:—(*a*) evaluating the net output of the various branches of productive enterprises, agriculture, mining industry, etc. It is necessary to avoid double counting e. g. counting both the output of wheat and labour of the cattle employed in raising it; (*b*) adding the value added to home produced goods and also to imports by transport and mercantile services; (*c*) adding excises on home-produced goods; (*d*) deducting the value of exports (f. o. b.) including gold and silver; (*e*) adding the value of imports (c. i. f.) including gold and silver; (*f*) adding custom duties on imports; (*g*) deducting the value of goods, whether home-produced or imported, which are used for the purposes of maintaining fixed capital or stocks or raw and finished goods intact: (*h*) adding the value of personal services of all kinds; (*i*) adding the annual rental value of houses, whether rented or lived in by the owners; (*j*) adding the increments in the holding of balances and securities abroad, whether by individuals or the government, or deducting the decrement in such holdings and or deducting the increment in the holdings of the balances and securities in the country by the residents abroad or adding the decrements of such holdings.

This method aims at evaluating the goods and services accruing to the people of India during a year but in the case of some of these processes it is necessary to bear in mind the following comments:—

In India a substantial portion of agricultural produce is consumed by the producer or bartered locally for services. It is necessary to value this portion of produce like the rest of the out-turn at its price at the point of production, not at the retail price which the consumers in distant markets pay, and which includes costs of handling, etc. which are not incurred on the home-consumed produce. It is necessary to add excises on home-produced goods and customs duties on imports as we are in search of the aggregate of exchange-values to the consumer. The operations in the case of (d), (e) and (j) can be easily justified. If a loan is raised in a foreign country the securities imported from the debtor country would form part of the real income of the investors. Again the increment of capital wealth, which is included in the evaluation of production or of imports, is balanced by a capital liability to foreigners and so must be deducted to find the net income. In the case of changes in ownership of bank balances at home and abroad and in increases and decreases in the stock of precious metals the above considerations are equally applicable. Strictly speaking, the services of the government servants should be included in the value of personal services of all kinds as they confer direct utility such as protection, amenity of life, etc. but such subtleties can be ignored to avoid complications at the initial stage. Besides this, in India the question of charitable gifts and of the payment for services of a customary or religious nature is particularly indefinite.

The Census of Products method is more fundamental of the two methods of estimating national income but with a view to tally the results of the Census of Incomes method with it the following precautions should be taken.

All self-consumed produce and receipts in kind must be included in the individual's income, valued at their selling price at the point of production. So must be the yearly value of the houses occupied by the owners. All interest payments, even on loans incurred for consumption purpose must be deducted before entering the individual income. The income of all individuals in the country including interest on government Securities should be entered *gross* i. e. before payment of direct taxes including land revenue. The income of government servants should be taken inclusive of pension rights accruing during the year. To this total should be added the undistributed profits of companies and the net profits of government enterprises. It is necessary to deduct from the total the amounts of interest on government loans other than for production work and the pensions of ex-government servants, whether due at home or abroad. To the total arrived at should be added receipts from customs and excise, stamp duty and local rates.

The authors of the report did not then recommend for India any estimate of national wealth as a whole. The two methods used elsewhere were of no use to India. The first is to capitalize the yield of all income-bearing property, including goodwill, assigning appropriate numbers of years' purchase to the rent of land and houses, dividends, interest, etc. and sometimes to add estimates of property publicly owned such as docks, railways government buildings, etc. The second is to use the statistics of property passing at death as shown by the probate and death taxes. This procedure is not possible in India as there are no taxes on inheritance. In the case of the first method too there is not enough information for ascertaining values.

Though under the circumstances it is not possible to measure with precision the total wealth it may be possible to give some indications of its change by series of estimates relating to public expenditure on permanent work, investment of new capital and to other expenditure of the nature of capital outlay.

The investigation proposed for estimating the national income is primarily on the basis of production but, as in similar estimates in all countries, a minor part depends on individual incomes. The proportion to be thus estimated is probably greater in towns in India. Partly owing to difference in the nature of the products and partly because different methods of investigation are necessary, rural income is distinguished from urban income. For rural income they recommend an estimate of the quality and value of all goods and services arising from the land or rendered in the village by the method of intensive surveys in selected villages. For urban income they advocate, in the first instance, surveys of the larger towns based on a sample enquiry of the personnel and occuptions of families, an estimate of their incomes by personal statements and by investigation of wages and salaries current in towns. Income tax statistics can afford valuable help for incomes over Rs. 2000. They have also recommended an intermediate urban population census. These three enquiries would be supplemented by a census of production applied to factories using power, mines and some other industries. There is a likelihood of overlapping but when all materials are assembled, this duplication can be estimated to a large extent.

Rural Surveys :—For the investigation of rural income the committee has recommended that an adequate number of villages should be selected by random sampling for conducting an intensive survey. A list of villages in a province should be made in geographical order of districts and they should be selected by starting from some random number and by keeping equal space between the villages taken. Once a village is marked for an investigation it should on no account be substituted by another. The report has given the minimum number of villages to be investigated in each province.

Province	Number of villages	Number in sample
Bengal	86,000	250
Bihar & Orissa	83,000	300
Bombay	21,000	200
Central Provinces	40,000	200
Madras	51,000	200
Punjab	35,000	200
U. P.	106,000	300

With a view to make the total for British India some estimates must be made for Assam, N. W. F. Province, Tea plantation of Bengal, districts of Bengal where coal-mining is important and the areas affected by earthquake and not fully resettled till the date of enquiry. Again the report has advocated that the investigators should be trained persons and they should establish personal contacts.

Urban Surveys:—In the case of urban areas surveys of larger towns is suggested in which enquiry should be made of the personnel and occupations of families. The problem is to be dealt with firstly by a synchronous survey of those cities having universities and secondly, by making similar surveys of other towns though not so intensively.

University City Surveys:—A central committee should prepare and outline of enquiry and co-operation should be sought of the economic staff under whose guidance, students of economics should carry on investigations. There are two methods of approach : occupational and by families. In the case of occupational census which is almost essential, enquiries should be made in each industry and important occupations in the town, regarding current rates of earnings and wages for the year making allowance for the seasonal variations. The enquiry should be extended even to clerks, municipal and railway employees, tonga drivers, etc. and not restricted only to those employed in constructive industries. It is necessary that the method of payment, piece or time wages, should be clearly mentioned. An accurate list of tenements should be made and in the case of big cities of more than 150,000 population they should be divided into wards or groups of which having units of about 30,000 houses. Of them about 1000 houses should be selected at random for investigation.

Census of Production:—The investigation should be made in the case of concerns employing 20 or more persons and using mechanical power but some small workshops, certain industries where mechanical power is not used such as brick-making, carpet manufacture, railway and all establishments under the Mines Act should be included. There should be a special act of the Central Legislature making it obligatory to supply facts when demanded and this work should be conducted by the Director of Statistics. The essential facts to be elicited are the aggregate value of the sales and the aggregate cost of materials for each factory. The classification of products should be the same as that of exports and imports. The employees should be classed as salaried persons and wage-earners, young and adult with an extra statement of the age division between the two sexes, male and female. Details of the amounts and values of different commodities produced and materials bought and of power used should also be obtained. In the beginning there will be difficulties and opposition to the investigators but with periodic repetition of the census they decrease automatically.

Government Publications:—The authors of the report have suggested the publication of a book of the type of the Guide to Current Official Statistics published in England. The purpose of this is to direct enquiries to all current official publications that contain statistics bearing on his subject and more especially to inform him of the nature of statistics he would find in the volume to which he is referred. This guide will give an idea of statistics already available and when a new enquiry is to be made it will be very useful. "The publications", the report says, "are out after a long interval after the period is over and great difficulty is experienced by the public in obtaining them." In fact an attempt should be made to make them popular as well as easily available.

The report has devoted one full chapter on serial statistics and has also discussed at length the formulation of different problems regarding population, natural resources, capital and organization.

Sources Of Statistics in India:—From the above historical account it will be seen that we have not got many sources of statistics in India. The main source, without doubt, comprises of the periodical reports and publications of the Government of India as well as of the Provinces and Indian States including those of the Municipalities and District Boards. The various committees and commissions such as the Royal Commission on Agriculture, Banking Enquiry Committee, Tariff Boards appointed by the Government have during the course of their enquiries, collected very useful statistics. The third important source is the collection made by semi-government bodies such as the Reserve Bank of India, the various railways and research departments of the different universities. But research agencies and research scholars given to the collection of statistics are comparatively very few in number. Trade and commercial organizations such as the various Chambers of Commerce and private organizations do attempt in this direction, but on a limited scale. Some economic and commercial periodicals such as *Eastern Economist, Capital, Commerce* maintain as well as publish at regular intervals trade statistics but it is mostly obtained from the Government. The work of the Imperial Council of Agricultural Research as well as that of the Indian Statistical Institute of Calcutta is of great importance in this direction. Private commercial concern as well as joint stock companies also collect some useful statistics to serve their purpose but the data are not published and very few have an access to them, as they are considered highly confidential lest the matter may by used by competitive concerns. In short, in India we have to depend mainly upon the Government or the semi-government institutions for statistical material.

Defects of Government Statistics:—In India, as already stated, we have statistics compiled and published mainly by the various government departments. Their collection, however, is attended with a number of difficulties. India is almost a continent and due to its vastness this work becomes difficult as well as expensive. The percentage of illiteracy is also very high and because of this, it is almost impossible to obtain active response and co-operation of the public in general. Schedules cannot be got filled by the informants and even if enumerators approach them they are likely to be suspicious because of their ignorance. Again, due to the existence of a few cities and more villages the population is widely scattered and for enumeration this is a great handicap. To approach people living in scattered areas means more time and energy as well as expense. There is also a great dearth of trained personnel to undertake efficiently statistical work. Because of these and other things statistics in India are defective in sevaral respects.

In the first place they suffer from lack of co-ordination. 'Statistics for India have largely originated as a by-product of administrative activities' and naturally, therefore, there was no planning for their collection and tabulation. Different departments of the Government collected them to serve their departmental purpose. The result has been that there is not only want of co-ordination but there has been in some cases duplication. Had there been an efficient organization at the centre as suggested by the Bowley–Robertson Commitee this could have been easily avoided. After the publication of their report, efforts are being made to organize this work properly but even now the progress in this respect is not quite satisfactory. The following remarks of Knibbs, Commonwealth Statistician of Australia made in January 1920, in his paper to the Royal Statistical Society aptly apply to India also. ' A Department whose duty it is to keep the Government, publicists and economists and th nation generally informed as to the movement of every importan activity in it, and of population facts, is of obvious value i intelligent direction is to be given to national affairs, or a intelligent study of them is to be made possible. To a larg extent, existing statistics are a side-product of various govern ment or other department, produced mainly as a sort of publi advertisement of their activity or for departmental use, eac acting on its own initiative, the whole un-coordinated and ofte without appropriate technical direction.

Secondly, Indian official statistics are not considered t be very accurate as well as reliable. The main reason for thi is that the agency employed for the collection of data is fa from satisfactory and this is much more so in the case agricultural as well as price statistics. Though governme officials try to present the statistics as very accurate, the dat collected cannot be said to be very reliable. In this connection th following extract from Sir Josiah Stamp's book, 'Some Econom Factors in modern Life' is very illuminating. Harold Co when a young man in India, quoted some Indian statistics t a judge, the judge replied, "Cox, when you are a bit olde you will not quote Indian statistics with that assurance. T Government are very keen on amassing statistics; they colle them, and add them, raise them to the nth. power, take the cu root and prepare wonderful diagrams. What you must nev forget is that every one of those figures comes in the first instan from *chowkidar* (village watchman), who just puts down what damn pleases".

The other reason for this want of accuracy is the use of old ante-dated methods of analysing the primary data. There also exists inconsistency as well as incompleteness in the statistics and this is mainly due to the statistics being the by-product of administrative activity as well as absence of planning and lack of co-ordination. If the collection of statistics is properly planned not only incompleteness will be removed but checks and counter-checks can be introduced. Another defect in the case of Indian statistics is their inadequacy. This has been clearly brought out by the remarks of the Indian Economic Enquiry Committee quoted at the beginning of this chapter. They are also not self-explanatory because prior to the publication of the 'Guide to Current Official Statistics' no effort was made to give their exact scope and significance. Above all, the most irritating defect is the inordinate delay in their publications. If they are made available in proper time they can serve as guide for future planning. For instance, the Review of Trade of India is out months after the official year is over. Even the trade figures about India's trade are published earlier by the British Trade Commissioner in India. Things can be improved to a great extent if an efficient staff is recruited for this work and proper organization and co-ordination is saught. Since the counry has achieved political independence, great scope for economic development is available and scarcity of proper statistical material is a great handicap to economic progress, as a country's planning has to be necessarily based on that and the earlier this organization is made the better.

Government Statistical Publications:—The various government offices collect and issue statistics relating to different topics such as prices, wages, agriculture, industry, trade and transport. A layman, however, used to experience a great difficulty in ascertaining where he would find statistics pertaining to his problem. In England by the publication of the booklet, 'Guide to Current Offical Statistics' this problem was solved long ago. The Bowley-Robertson Committee recommended that such a guide should be published in India.

Guide to Current Official Statistics:—This guide was first prepared in November, 1941 under the instructions from the

Economic Adviser for India. Its first volume comprises the following sections:—

I. (a) **Primary Production :** (i) Agriculture, (ii) Animal husbandry including fisheries, (iii) Irrigation, (iv) Forests, (v) Minerals.

(b) **Industrial Production :** (i) Large-scale manufacturing industries, (ii) Power production, (iii) Cottage industries.

II. **Prices.**

The second volume, which was got ready two and half years later includes the following sections:—

I. **Trade** (a) Foreign Sea-borne Trade, (b) Coastal Trade, (c) Land Frontier Trade, (d) Inland Trade and (e) Miscellaneous.

II. **Transport and Communications:**—(a) Railways (b) Roads, (c) Shipping, (d) Canals, (e) Civil Aviation, (f) Posts Telegraphs and Telephones and (g) Radio and Wireless.

III. (a) **Joint Stock Companies and** (b) **Index of Profits.**

IV. **Insurance.**

V. **Co-operative Societies.**

VI. **Banking.**

VII. **Currency and Coinage Exchange and Bullion** (including securities).

In each of these volumes under different headings is given the list of current official publications on the particular subject and short description has been given of the matter to be found in the publications. At the end an index is appended. For convenience of ready reference the index has been divided into two parts. The first is the 'Publications Index' in which under the head of each of the publications arranged alphabetically a detailed reference to all the items included in the guide has been given in alphabetical order. The second is the 'Subject Index' which the different items dealt with in the guide have been alphabetically arranged and under each item the various publications relevant in the context have been referred to in alphabetical order.

As there are several official publications giving statistical matter it is not possible to examine all of them in this small volume. Effort has, therefore, been made to give in brief the important statistical matter contained therein under different topics.

Agricultural Statistics:—Even to-day India is dependent largely on its agriculture and so it is needless to emphasise the importance of reliable and detailed statistics regarding agriculture. Land revenue has been one of the principal sources of income to the Indian Treasury and so we possess statistical records collected by revenue officials in the course of their administrative work. Even as early as 1866 we find agriculture statistics compiled by provincial governments. The first attempt to forecast wheat crop can be traced to 1885.

Statistics relating to agriculture are available in the following important current official publications :—

1. *Agricultural Statistics of India, Volumes I and II* :— They are issued annually based mainly on the annual returns furnished by the Provincial Governments and Administrations to the Department of Commercial Intelligence and Statistics. In the first volume are given statistics relating to British India regarding area, prices and land tenure and revenue. Details are given about the area according to the professional survey, area under forests, area not available for cultivation, cultural waste other than fallow, current fallows and net area sown. In the section on the price are given the prices of important crops at harvest time. The number of crops for which prices are available varies from four in Madras to twelve in Bihar and Orissa. The second volume deals with the statistics relating to Indian States. These volumes also give details about area and irrigation with individual figures for rice, wheat, barley, jowar, bajra, maize, sugarcane and cotton in addition to other minor crops.

2. *Estimates of Area and Yield of Principal Crops in India.* This is published annually and the crops dealt with therein

are divided in three groups, (a) Rice, wheat, sugarcane, cotton, jute, linseed, rape and mustard, sesamum, castorseed and groundnut, (b) Tea, coffee and rubber and (*c*) Barley, jowar, bajra, maize, gram, tobacco and indigo.

3. *Plantation Statistics:*—comprise of area, production, average number of persons employed daily and exports regarding tea, coffee and rubber.

4. *Statistical Abstract for British India.* It gives in addition to the area under important crops and the yield as given in the Agricultural Statistics of India volume I, figures of rainfall and temperature.

5. *Quinquennial Report on the Average Yield per Acre of Principal Crops in India.*

Besides these publications periodical crop forecasts are issued from time to time about cotton, wheat, sesamum, rice, jute, castorseed, etc. *Agriculture and Animal Husbandry*, issued annually by the Imperial Council of Agriculture Research is a very valuable publication as it gives particularly the progress of research on the various crops and the improvements effected in the method of production. *Indian Trade Journal* published weekly and *Monthly Survey of Business Conditions* also contain useful agricultural statistics.

The variety or agricultural statistics is very often questioned mainly because of the type of agency employed in collecting them. In India the work of making primary estimates is done by village accountants, who neither possess the requisite training nor have adequate time. They are saddled with the Revenue collecting work besides other administrative duties and naturally, therefore, very often the collection of statistics is not attended to by them but by *chaukidars* who are not quite literate. Many a time the returns are filled by looking to the old record, with a few alterations based on general enquiry or information or even by making some changes just with a view to show that they are accurate. These figures reach the provincial government through the District and Taluka officers who also make snch changes as they deem necessary, not in any case based on further satatistical investigations. All India statistics are compiled from the figures received from the provinces. The village accountants usually report better conditions both as regards the cultivation as well as crop and its yield. This is done to avoid giving much remission in the payment of land revenue to the state. They do not report great changes as in that case they will have to be explained. Forecast made on *Annawari* system are far from satisfactory. These estimates being generally in the form of so many *annas* by taking sixteen annas as standard. An error of half an anna is very likely to be perpetrated and its effect in the cases of a crop below the normal will be large in the case of eight anna crop; it being more than 6%. This method should be changed to that of stating the yield in maunds per bigha or in percentages. In the case of yield too it should be ascertained by means of crop cutting experiments based on random sampling. Even the fixing of standard yield is fairly complicated as it does change with the lapse of time as well as with the introduction of changes in agricultural methods. It is very necessary that the staff engaged in this work should be properly trained. Sometimes there is no uniformity in the computation of cultivated area particularly in the case of land yielding more than one crop.

In places where Ryotwari or Temporary settlement system is in vogue statistics collected are more accurate than those from the permanently settled areas such as those in Bengal and Bihar. This is chiefly due to the absence of village accountants and subordinate revenue officials. The estimates in such areas are prepared by village headmen through the help of the police and as such. they are not very reliable.

In the case of livestock the census taken is quinquennial. There is no proper statistical data regarding animal products such as meat, milk, eggs., hides, bones. Because of some of the states not joining the quinquennial census of livestock the figures presented are likely to be not very accurate. The present classification also needs to be widened. Besides, owing to the lack of uniformity in the time and method of enumeration, it is believed that the reported figures are not free from errors.

Transport and Communication Statistics :—Transport statistics can be classed under five heads :—(A) Railways,

(B) Roads, (C) Shipping, (D) Canals and (E) Civil Aviation. Excepting the figures relating to country carts, all others are collected by departments in direct touch with the respective branches and so they are fairly correct and reliable.

(A) **Railways**:—Statistics relating to railways are available from the following publications:—*Annual Report of the Railway Board on Indian Railways, Volume I and II.* These volumes furnish information relating to the mileage of railways, the result of their working regarding the number of passengers and the tonnage of goods carried and the earnings therefrom. Besides these the financial results showing traffic receipts, working expenses, density of traffic, etc. are also given. The first volume gives a description of the general position drawing attention to the important features noticeable in the Statistical. Tables which are to be found in volume. II. (2) *Statistical Abstract for British India*—In the publication all the important tables from the Annual Reports of the Railway Board are reproduced giving figures for the years in every case. Besides these, statistics relating to earnings and wagon loadings are available from the *Indian Trade Journal* and *Monthly Survey of Business Conditions.*

(B) **Roads**:—Information relating to road transport is mainly available from Statistical Abstract for British India. It gives length of roads mentioned by public authorities as well as the number of motor vehicles running in British India. '*Indian Roads*' issued monthly by the Departments of Communications gives statistics of extra-municipal roads mentioned by public authorities and also contain details concerning the nature of the roads i. e. whether bituminous cement, macadam, etc. Besides these both the volumes of Agricultural Statistics give useful statistics about the number of carts in the various provinces based on periodical censuses held in the various provinces and states.

(C) **Shipping**:—One of the main sources for statistics on this branch of communication is '*Accounts relating to the Sea-borne Trade and Navigation of British India* issued by the Department of Commercial Intelligence and Statistics. It gives details about the movements of cargo shipping, such as the number and tonnage of vessels that are entered and cleared with cargoes.

Account relating to the coasting trade is also published monthly and contains details regarding coastal voyages. *The Annual Statement of the Sea-borne Trade of British India. volume* II contains particulars of the number and tonnage of vessels that entered into and cleared from British Indian Ports classified according to the countries whence the ships arrived and to which they cleared, the nationality of the vessels, and the port into or from which the ships entered or cleared. It also gives the number and tonnage of vessels built in the several ports of each province as well as the number and tonnage of vessels first registered in British India. *The Statistical Abstract for British India* is also very useful as it gives details regarding cargo traffic, the number of vessels belonging to or registered in British India, returns of wrecks and casualties in Indian waters, and passenger traffic. *Monthly Survey of Business Conditions* also contains monthly figures of tonnage entered or cleared at British Indian Ports.

(D) **Canals** :—Canal transport in India is not much developed. The *Annual Statistical Abstract for British India* gives details concerning the length of canals open for navigation total number of boats plying cargo and the quantity of cargo carried and the total number of boats and that of passengers carried. *Annual Review of Irrigation in India and Triennial Review of Irrigation in India* both issued by the Government of India contain figures of the length of navigable canals available for use in the various province.

(E) **Civil Aviation**:—The only official publication furnishing civil aviation statistics in the *Report on the Progress of Civil Aviation in India* issued by the Directorate of Civil Aviation. It gives particulars regarding air mails carried, personnel employed number of flights, number of passengers, total freight carried and also operational statistics relating to the various Indian air services.

Communications :—Statistics regarding communications other than transport are classified under two heads ;—(A) Posts, Telegraphs and Telephones and (B) Radio and Wireless Statistics relating to the former group are available in (1) *Annual Report of the Post and Telegraph Department* (2) *Statistical Abstract for the British India* and (3) *Monthly survey* *of Business Conditions.* Statistics relating to Radio and Wireless are also found in the *Annual Report of the Post and Telegraph Department and the Statistical Abstract for British India. The Report on the Progress of Broadcasting in India* issued by the controller of Broadcasting gives figures of the number of broadcast receiving licenses and that of programme transmission hours.

Industrial Statistics:—In India Industrial establishments employing not less than twenty persons per day are included in large industrial establishments and statistics relating to these are available but not for cottage industries except in the case of the handloom industry.

Statistical information relating to large-scale industries is available from the following publications:—

(1) *Large Industrial Establishments in India:*—It is a biennial publication and contains statistics of average daily number of workers and also of factories which are classified as (a) Textiles (b) Engineering (c) Minerals and Metals (d) Food, drink and tobacco (e) Chemicals and Dyes (f) Paper and Printing (g) Process relating to wood, stone and glass (b) Process connected with hides and skins (i) Gins and presses and (j) Miscellaneous included therein are grass farms, mints, ordnance factories, rope-works and rubber works. Perennial and seasonal factories are considered separately. The latter group consists mostly of those dealing with food, drink, and tobacco, and gin presses.

(2) *Monthly Statistics of Cotton Spinning and Weaving in Indian Mills:*—This give statistics of quantity of yarn spun in the mills and quantity of manufactured goods. The figures are compiled from returns submitted by the mills and are given in separate tables for each province and for all other territories.

(3) *Monthly Statistics of Production of certain selected Industrials of India:*—The publication, which is issued six times in a year, contains statistics of (a) Jute manufactures, (b) Paper manufactures, (c) Iron and Steel manufactures, (d) Petrol (e) Kerosene oil, (f) Cement, (g) Wheat flour, (h) Paints, (i) Heavy Chemicals, (j) Sugar, (k) Matches and (1) Products of distilleries and breweries.

(4) *Indian Trade Journal*:—This weekly journal furnishes information regarding the production of refined sugar, mill consumption of Indian cotton and the amount of cotton pressed in the cotton pressing factories in British India and in the Indian States.

(5) *Statistical Abstract*:—This important publication supplies information regarding the number of factories in British India with the average daily number of workers classified according to sex as well as into three broad groups; (a) adults who have completed the seventeenth year (b) adolescents, who have completed the fifteenth year but not the seventeenth year and (c) children who have not completed the fifteenth year. It also gives the number of factories in Indian states and the average daily number as well as production of yarn and piecegoods in cotton mills and the production of certain selected industries as are found in the Monthly Statistics of Production of Certain Selected Industries in India.

(6) *Statistics of Factories*:—It is issued annually by the Labour Departmest of the Government of India and gives the distribution of the number of workers in cotton and jute mills and the classification of factories according to the length of the intervals granted and the normal number of working hours per week. It also gives details relating to the number of convictions for branches of the Factory Act.

(7) *Monthly Survey of Business Conditions*:—It is issued by the office of the Economic Adviser and gives statistics regarding production of cotton piecegoods and production of jute, iron and steel and sugar manufactures. Summary tables comparing two years' production figures and charts showing the variation over a number of years in the production of the above industries are given. The publication also provides monthly statistics of electrical power, generated and consumed.

Cottage Industries:—As already remarked the only cottage industry for which all-India estimates of production have been attempted from time to time is the handloom industry and they are found in the Review of Trade of India. The Census Reports and the report of the Indian Tariff Board regarding the grant of protection to the Indian Cotton Textile Industry contain statistics relating to the handloom industry. It is very necessary for the Government to collect statistics of soap as well as of vegetable products industries as both of them have attained considerable importance.

Price Statistics:—The price statistics can be classified under two heads, *wholesale* and *retail*. Wholesale prices of all the principal commodities entering the trade of the country are based on the reports from commercial bodies at important commercial centres and so they are to a large extent satisfactory. The quotations for retail prices are mainly for foodgrains and salt, but they are not very satisfactory as they are compiled from the reports of revenue officials. Because of their preoccupation as well as want of adequate training the reports are not very correct. Sometimes they are sent without sufficient personal enquiry and also by not taking care to see that the quotations are for the same quality of the commodity. It has been noticed that even though the market prices have changed appreciably the prices statistics do not record the change to that extent.

Statistics relating to prices are available from the following important publications:—

(1) *Indian Trade Journal*;—This weekly publication by the Department of Commercial Intelligence and Statistics gives statistical information regarding the wholesale prices of certain selected articles at various centres in India. These figures are published once a month and represent prices prevailing at the end of the month. It also publishes the monthly index numbers of wholesale prices in Calcutta, weekly price quotations of important commodities in selected markets, retail price of salt, harvest prices of certain principal crops, monthly figures of prices paid for sugarcane delivered at the factories and the price quotations for tea at auction sales in Calcutta.

(2) *Index Number of Indian Prices:*—This quenquennial publication gives available information relating to prices of exported as well as imported articles, Index Numbers of retail prices of foodgrains and weighted Index Numbers of wholesale prices. For this index number important commodities like

groundnuts, pig iron and steel manufactures are not taken and so it is not safe to rely on this Index Number as a sure guide to price movements,

(3) *Statistical Abstract far British India* :—This publication gives monthly Index Number of wholesale prices in Calcutta, Annual Index Numbers of prices in India for exported, imported and all articles, Annual Index Numbers of retail prices of foodgrains in India. Wholesale prices of certain staple article of trade at selected stations, monthly Index Numbers of working class cost of living in certain industrial towns, average annual retail prices of sales and average annual retail prices of foodgrains.

Besides these main sources, price statistics are also available from wholesale Prices of certain Staple Articles of Trade at Selected Stations in India, Agricultural Statistics of India, Monthly Survey of Business Conditions, Review of Trade of India and Weekly Market Reports relating to Wheat, Linseed and Rice. Most of the information contained in these books is found in the publications mentioned already.

Cost of Living Statistics:—In the chapter on Index Numbers details have been given regarding the principal Index Numbers in the country. Monthly Survey of Business Conditions in India and Statistical Abstract for British India are the two important publications for obtaining statistics on this topic. The former gives monthly index number of working class cost of living in some important cities. They are not, however, directly comparable, particularly as the based periods differ and some ignore the important item like house-rent. Besides. various provinces issue Index Numbers which would be of great help in the study of this important topic. It is however, necessary that index numbers should also be computed for middle class families also.

Wages Statistics:—In India wages statistics are not available to the extent they should be and it is much more so in the case of rural areas. In many cases labourers are paid in cash as well as in kind and are also given some other concession. Again in rural areas and even in towns it is not easy to collect reliable data for want of trained staff due to the absence of active co-operation on the part of the labourers. Statistics of wages paid in factories and mines are available to a large extent from the returns received from factories. For other urban occupations such statistics are not normally available. In the case of rural occupations customary statements supplied by revenue officers are available but for reasons already stated they cannot be completely relied upon. This problem is of great national importance and needs immediate attention on the part of the Government as well as of social workers.

Trade Statistics:—In India official trade statistics arise out of compilations made in the course of administration of laws such as those relating to taxation of imports and exports and from compilations made from returns received from the railways, the latter really covering only a small section of India's inland trade. The external trade includes sea-borne trade with foreign countries as well as across land frontiers, and the internal trade comprises of rail and river-borne trade as well as coastal trade. The important official publications furnishing detailed trade statistics are (1) Accounts relating to the Sea-borne Trade and Navigation of British India issued monthly, (2) Annual Statement of the Sea-borne Trade of British India with the British Empire and foreign countries issued in two volumes, (3) Accounts relating to the Sea-borne Trade and Navigation of British India for the calender year published annually, (4) Trade Statistics relating to the Maritime States in Kathiawar and the State of Travancore (Monthly), (5) Accounts relating to the Coasting Trade and Navigation of British India (Monthly), (6) Trade at Stations adjacent to Land Frontier Routes (Monthly), (7) Accounts relating to the Inland (Rail and River-borne) Trade of India (Monthly), (8) Exports of Indian Artware and Sports Goods (Monthly), (9) Review of Trade of India published annually and (10) The Indian Trade Journal (Weekly). Besides these much useful information regarding trade statistics is available from the Monthly Survey of Business Conditions in India issued by the Economic Adviser and the Statistical Abstract for British India published annually.

Of these the Annual Review of the Trade in India gives comprehensive information regarding foreign trade of India including the maritime states. It is divided into two parts : the Report and the Statistical Tables. The reports contains observations on general trade developments inside the country as well as abroad; account of the chief influences at work in the field of trade, particularly in the markets for the staples of India's foreign trade and on conditions prevailing in the fields of agriculture, industry, and economic policy in so far as they influence the course of trade. In this volume is also given an analysis of particular items of imports and exports trade as well as directions of overseas trade. Various statistical tables are given showing analysis of trade figures according to countries as well as commodities. Information is also available from this volume regarding land-frontier trade. In short, it is a very useful publication, only that it is published long after the completion of the year and imports and exports on the government account are not given in as much detail as those on private account. This comes in the way of finding out accurately the total trade in particular commodities.

National Income:—According to Dr. Bowley 'By Total National Income' is generally meant the aggregate of the incomes (including earnings) of the persons comprising the nation; income taken as meaning the money value of goods coming into a person's possession during a year for his own use (subject to a rate and taxes) after all expenses connected with obtaining it are subtracted.' There are three main methods of measuring it: (a) Census of production, (b) Census of Income and (c) Occupational Census. Unfortunately for want of reliable and complete statistical data even one of these methods cannot be applied to obtain satisfactory results. The suggestions and remarks of the Bowley-Robertson Committee on this topic have already been given at length.

Inspite of the want of complete as well as very reliable data attempts have been made to estimate per capita income by many persons the results of some of them are as under:—

Estimates of *per capita* Income for India

Author		Year for which estimate was made	Income Rs.
Dadabhai Navroji	...	1868	20
Cromer and Barbour	...	1882	27
F. G. Atkinson	...	1895	35.2
Lord Curzon	...	1897–98	30
William Digby	...	1899	18
Wadia and Joshi	...	1913–14	44.3
Shah and Khambatta	...	1921	74
Findlay Shirras	...	1922	116
Vakil and Muranjan	...	1925	74
V. K. R. V. Rao	...	1931–32	65

The estimated figures in the above table vary considerably and so in comparing the per capita income several points must be borne in mind. In the first place, these figures relate to different years and so it is very necessary to assess carefully the economic condition of the people as between the different years. Due allowance must be made for the change of the Index Numbers of prices in these years. Secondly, it must be seen whether the area covered by the computators is the same or not. Most of the authors have prepared their estimates for British India but the calculations of some of them, including Prof. Shah and Khambatta, are for the whole of India including native states. Again different methods have been used by different persons. Principal Findlay Shirras has not made any deductions while others have made them but in varying degrees. It is, however, hoped that after the statistical department has been properly organized it would not be difficult to have a reliable data for ascertaining the National Income.

The Census Reports:—It is customary on the part of the Government to arrange for a decennial census for the collection of social, political religious and educational facts concerning the population of its country and to issue Census Reports which are of great value to its administrators and also to economists, sociologists and businessmen. With the help of census reports the potential military power of the country can be easily estimated. For improving the sanitation and introducing measures to improve the birth and death rates the necessary data can be easily culled from them. These rates pertaining to the various countries can with profit be subjected to a comparative study. The problem of growing population and food supply can be studied by referring to census figures and so also to the changes in the social habits, literacy, age at marriage, etc. The question of labour supply and markets in different areas of the country can be examined with advantage by industrialists for starting new concerns and opening sale depots. Businessmen will find ample material to calculate per capita consumption of the various articles and thereby to push their goods. Firms interested in transport industry can get a very useful data for the expansion of transport facilities. In India, particular importance was of late attached to the census figures because of the government's policy of reserving seats in the legislature and in the various public services by taking population of the different communities into account.

In India the custom of making periodic estimates of population was prevalent in several provinces but the first attempt to take a general census was made between the years 1867 to 1872. Even then many native states were left out and the enumeration was nun-synchronous. The first regular census on modern system was carried out on 17th February, 1881. The count was synchronous except in some remote and jungly tracts where the enumerators were unable to move at night. The second general census was taken on 26th February, 1891 and the arrangements were more efficient. The third and the fourth census were taken on 1st March, 1901 and 10th March, 1911 respectively. This was followed by another count on 18th March, 1921. The sixth census was taken on 11th January, 1931 but due to distributed political atmosphere and more so because of the census boycott movement the figures recorded cannot be said to be very accurate. The census of 1941 was a very different phenomenon as it happened to be non-synchronous having 1st March, 1941 as the Central Day. The enumerators were working for three days but the entries recorded were as they would have been on 1st March, 1941.

In India there is no permanent establishment for census work. The Census Commissioner for India is appointed about a year before the actual census time and under him work the various Provincial as well as State Superintendents of census operations. Each Province is divided into Districts, Charges, Circles and Blocks. Wherever there are municipalities they usually arrange for the census of their areas. The data of the census is kept as near as the decennial anniversary of the previous census and occasions of large fairs or other gatherings which disturb the normal distribution of population are as far as possible avoided. Before 1941 the census used to be taken at the same time all over the country and the counting was based on *de facto* principle i. e. people were taken into consideration where and when the census was taken. In 1941 the one-night theory was replaced by a period system and the usual place of residence was taken into account on the *de jur* principle. Prior to 1941, slips were prepared and tables were got ready therefrom. Now the enumerators directly make entries on slips, thus effecting considerable economy.

The main changes, introduced in the operations of the 1941 census, or *truncated* census as it is called, were many. Apart from the abolition of one-night theory of enumeration and the writing of schedules thus saving the labour and charges of copying slips, the forms prepared had only numbers given for each item and one explanatory sheet was given for each pad in the language of the area. This meant a considerable saving in printing charges as formerly forms were required to be printed in several languages prevalent in the country. Except in few provinces the caste column was done away with and enquiries were made regarding the age of the woman at the time of the birth of her first child. Enquiries were made regarding communities and not religion.

The census work in India is beset with a number of difficulties. Due to illiteracy schedules cannot be got filled by the people and even competent enumerators are difficult to be had particularly as the work is honorary. Due to the employment of a large number of hands mistakes occur for want of intensive training. Charges in the territory of the country e. g. separation of Aden, Burma, Pakistan, etc. have made comparisons of figures with those of the past very difficult. It is mainly because of the honorary services that the financial difficulty is not experienced. The American census for a third of our population costs 12 crores that is 24 times the cost of the Indian census. The cost of census in England is Rs. 150 per thousand while the Indian figure comes to under Rs. 15.

For reasons of economy the Government of India did not sanction full grant of expenses in 1941 with the result that although the census had been carried out on elaborate scale the results were not tabulated except for the variation of population. Tables giving such important information as age, means of livelihood, or literacy were shelved except for some sample surveys.

If we critically study the census reports as well as the conditions under which they are compiled we must admit that due to want of a fair proportion of trained as well as permanent staff there is a marked lack of uniformity in the classification of occupations. Due to illiteracy as well as unwillingness to give the correct age on the part of unmarried women, age returns are not what they should be. As there is no definite standard test of literacy the decision regarding whether a person is literate or illiterate is taken differently by different enumerators. Again people have a tendency to avoid telling the correct cause of death. This is much more so in cases of infectious diseases, deafness, insanity and blindness are also concealed as far as possible.

As already stated in 1941 the Government did not sanction enough money for tabulating complete details. In general however, from the census reports we get information regarding population, its distribution and movement, urban or rural, birth place, migration, age, sex, civil condition, infirmities, occupation, literacy, languages, religion caste, etc. It is however, hoped that with growing literacy people will be more census-minded and it will make easy in time to have more accurate information.

Intercensual Estimates of Population:—There is a decennial census and it is a costly as well as lengthy process to have it often different methods are used to have intercensual estimates of population. The easy method of arriving at an intercensual estimates of population is to add to the previous census figure the number of births and immigrants since the date of the last census to the date of the estimate and to deduct the number of deaths and emigrants in the same period. The correctness of the result will depend upon the proper recording of births and deaths as well as migration statistics.

The second method, often known as the A. P. Method is based upon the assumption that the population increases in the arithmetic progression. The rate of increase per year is ascertained by using the arithmetic progression formula and to the previous census figure is added the likely increase for the number of years that have elapsed since the previous census. This method is defective as it is based on the assumption that the rise is equal for all years and does not take into account the spread of epidemics and outbreak of wars both of which are responsihle for a considerable loss of life.

The third method, better known as the G. P. method, is based on the assumption that a rise in population is in geometric progression and not in arithmetic progression. The process followed is identical as in the previous case only that the years are multiplied by the rate ascertained by geometric progression. This method like the previous one fails to take into account the effect of abnormal factors like pestilence and war.

As far as India is concerned both the A. P. and G. P. methods are not suitable as there is very often an outbreak of epidemics and a large toll of life is taken by them. For using the first method the difficulty is of getting correct statistics but on the whole that is the most suitable method of arriving at intercensual estimates of population.

Vital Statistics:—This is one of the most important branches of statistics dealing with the mankind in aggregate. Broadly speaking, it includes the whole study of man as affected either by hereditary or by environment so far as the results of this study can be arithmetically stated. The main items of vital statistics, therefore, are number, age, sex. mental condition, occupation, marriages, births, deaths, migration and sickness. For the information on these topics we have to depend mainly on the census reports. In short, it is the science of numbers applied to the life history of communities and nations. It is also called 'Bio-statistics.'

The importance of an accurate study of vital statistics is very great, as thereby we can ascertain and classify facts for preparing our way to investigate causes of the various evils and to find ways of improving them. The high death rates particularly due to infantile mortality and maternal mortality can be improved. The infantile mortality rate in India is at least three times that of leading western countries. Maternal mortality is calculated by experts at 20 per 1000 in India while that in England is 2.9 only. It is estimated that even 'if one percent reduction can be effected it will mean 63000 women saved every year and roughly six lakhs in ten years with an addition of children in the case of women of child-bearing age. Apart from the aid of scientific investigation and improvement in the conditions of life, vital statistics are of great service for acturial work. Insurance companies base their premium tables on them. Even for legal purposes such as birth and death certificates they are of great importance. Even obtaining knowledge of civil conditions and their sources in itself is of great value to us. According to Whipple 'it is of the greatest importance to a nation that accurate records be kept of its vital capital, of its gains by birth and immigration and of its losses by deaths and emigration, for a nation's true wealth lies not in its lands, not in its forests and mines, not in its flocks and herds, not in its dollars but in its healthy and happy men, women and children'.

In India according to the Registration of Births and Deaths Act of 1886 this record is made in villages by village headmen and in the cities by municipalities. Unfortunately, however, this work is not properly done for various reasons. The illiteracy is to a great extent responsible for this state of affairs. People do not realise adequately the importance of this work. The Government too has not taken proper care to have an efficient machinery for this work. The village headman is ill-paid and lacks the requisite training. It is not usual on their part to wait for recording the births for a couple of days so that in case the baby dies he would avoid entries of births as well as of deaths. The causes of death are not scrutinised and particularly diseases of contagious nature are not mentioned correctly. Again, in all the provinces of the country there is no uniform system. This state of affairs needs immediate and thorough reforms. It is very necessary that the Government of India should control this branch and establish uniformity in the work done all over the provinces. The present Birth and Death Reports should be improved and checks and counterchecks introduced. It is very essential that proper record is made of the cause of deaths by laying down definite instructions in the matter. All this can be facilitated by impressing upon the people the unity of the record of vital statistics.

Exercises

1. Describe briefly the procedure followed in taking the Census of Population, in India. What are the main difficulties to be overcome in the census operations. (1948–Mar.)
2. Give a critical account of the available official statistic relating to vital statistics and yield of food crops in India. (1948–Mar.)
3. Examine the problem of estimating inter-censual population in large cities. (1947—Oct.)
4. Write a critical note on the official estimates of area and yield of the principal crops in India. (1946–Oct.)
5. Describe briefly the system followed for taking the census of population in India. In what respects did the procedure followed at the 1941 census differ from earlier practice. (1947–Mar.)

6. Discuss the comprehensiveness and reliability of Indian Agricultural statistics. What steps have been taken recently to improve their efficiency ?

7. Give a short account of the nature of information continued in one of the following :—(1) Annual Statement of the Sea-Borne Trade of British India, vol. I (2) Agricultural Statistics of India, vol I and vol. II (1944–Oct.)

8. Examine the adequacy and accuracy of either Agricultural Statistics or Price Statistics in India. (1944–Mar.)

9. Describe briefly the existing methods of estimating yield of agricultural crops in India and suggest improvements. (1943–Oct.)

10. What do you understand by the 'National Income of India' ? State briefly the material available for its estimation. (1943–Mar.)

11. Describe briefly mentioning sources what statistical material is available in India relating to any *two* of the following:—(i) Acreage and yield of important crops, (ii) Imports and Exports, (iii) Internal Trade and (iv) Vital statistics. (1943–Mar.)

12. What are the sources from which statistical data relating to large scale industries in India can be obtained ? Suggest improvements. (1942–Mar.)

13. Explain the importance of vital statistics to a country. Point out defects in the system of recording vital statistics in India and offer suggestions for their improvement. (1938–Mar.)

14. Give the main recommendations of the Bowley–Robertson Committee regarding the organization of the Department of Statistics in India.

15. Mention the important sources from which you get price statistics.

16. What statistical information will you be able to have from the following :—

    (*a*) Indian Trade Journal.

    (*b*) Agricultural Statistics of India.

    (*c*) Monthly Survey of Business Conditions.

APPENDIX I

SPECIMEN OF BLANK FORMS AND QUESTIONNAIRES

A few more blank forms and questionnaires have been given in this appendix to serve as specimens and they are as under :—

(1) Questionnaire issued by the Bombay Economic and Industrial Survey Committee.

(2) Form issued by the Labour Office, Bombay to collect information regarding workmen's family income and expenditure.

(3) Form of enquiry issued by the School of Economics and Sociology of the University of Bombay for an enquiry into the effects of inflation and price control on middle class families in Bombay City.

(1) Questionnaire issued by the Bombay Economic and Industrial Survey Committee

1. Name and address :—
   Established in—
   Industry.
2. Manufacturers of :—

Kind of Products	Quantity produced in a year	Value of production in a year	Remarks

3. Number employed :—
   Manual staff ... ...
   Clerical staff ... ...
   Officers ... ...
   Technical experts ... ...

4. Do you obtain any of your materials from abroad ? If so, is it possible to replace them by Indian substitutes ? If not, why not ? ...

5. What is the amount of capital invested in your concern ? ...

6. (*i*) Is your equipment efficient? In what direction is improvement possible in the same ? Do you think the State can help you in securing such improvements ? If so, by what means ? ...

(*ii*) Do you meet with any difficulties in securing equipment? If so what are they ? How do you think they can be removed ? ...

(*iii*) Do you use power? If so, is it steam, gas, or electricity? What is your annual expenditure on power?

7. How far, in your opinion, do your products compete with the products of Indian cottage industries or of small-scale factories ? ...

8. Is your industry seasonal? If so, why? ...

9. What is the usual market for your product? Local, district, provincial, inter-provincial, or foreign ? ...

10. Have you any difficulties in marketing your products ? How can they be removed ? Do you want any Government assistance in this matter ? If so, of what kind ? ...

11. Do you deal only with wholesalers or do you also organize retail sales?

12. Do you find any difficulties in obtaining capital ? If so, describe them. How do you suggest, they can be removed ?

13. In what types of your production do you meet with competition ? Please indicate the sources of campetition, whether from Bombay Province or the rest of India or from abroad...

14. What steps do you think can enable you to withstand the competition successfully ? Indicate the kind of help which you think you should get from (*a*) the Department of Industries, (*b*) the Co-operative Department, (*c*) other departments of Government ? ...

15. What is your total annual wage bill? Please give separately your wage-total for :— ...

    (*a*) manual workers ...
    (*b*) clerical staff ...
    (*c*) officers ... ...
    (*d*) technical experts ...

16. Have you any complaint regarding the efficiency of your workmen ? If you have, can you suggest any remedies ? ... ...

17. What, do you think, are the difficulties in the organizations of your industry and what measures do you suggest to remove them ?...

18. What, in your opinion, are the future prospects of your industry? Have you anything to say about the possibility of promoting the expansion of your industry? ...

19. Can you suggest the establishment of any auxiliary or subsidiary industry, arising from your industry?...
If there are possibilities of doing so, why have they not been established?

20. What, in your opinion, are the steps that the Government should take to facilitate such establishment?

21. Has the Government of Bombay taken any action during the last 17 years which has resulted in encouraging or promoting the Development of your industry? ...

    If so, please describe the steps taken by them in this connection?

22. Can you suggest any specific ways in which Government can assist in the development of your industry? or of subsidiary or auxiliary industries?

(2) Workmen's Family Income and Expenditure

Ward____________

Street____________

Tenament No____________

Reference No.____________

Notes of Scrutiny

(1) Is the head of the family literate ?____________

Place of birth : (District)________ (Province)________

Family Income and Expenditure for month of____________ 19

Name and address of the head of family____________

Religion and caste of family____________

Number of ages of persons in family

—	Men	Women	Children under 15 years		Relationship of each to the head of the family and age.
			Male	Female	
Number of persons residing at above address					
Dependants not living at above address					
Number of boarders ...					
Number of lodgers					

Family income of those residing at above address during the month of 19 .

Wage-earners				Earnings in above month												
—	Relationship to the head of the family	Age	Industry and Occupation (1)	Wages including overtime pay			Bonus*			Additional income				Total		
										Amount			Source			
				Rs.	a.	p.	Rs.	a.	p.	Rs.	a.	p.		Rs.	a.	p.
Men																
Women																
Children ... Male ...																
Female.																
Total amount of family income 2																

*Details of concessions given by the employer

Expenditure on food, fuel, lighting and rent

COMMODITIES (1) (*Note*:—This list may be altered according to what is purchased)	Total quantity consumed in month	Total cost in month
Rice ...		Rs. a. p.
Wheat		
Wheat flour		
Jowar		
Bajra		
Other cereals		
Turdal		
Gram		
Gramdal		
Mugdal		
Other pulses		
Raw sugar (gul)		
Sugar (refined) ...		
Tea (Powder or dust) ..		
Sweetmeats (consumed at home)		
Tamarind and kokam ...		Rs. a. p.
Turmeric		
Cocoanut and copra ...		
Other condiments and spices...		
Potatoes		
Onions		
Vegetables		
Fruits		
Pickles and preserves ...		
Ready-made tea		
Refreshments bought and consumed away from home ...		
Other food		
Total expenditure ..		

Fish ...					
Mutton ...					
Beef ..			Charcoal ...		
Eggs ...			Soft Coke ...		
Fowls ...			Firewood ...		
Other meat ...			Kerosene ...		
Milk ..			Caster oil ...		
Curds and butter milk ...			Match boxes ...		
Ghee ...			Candles ...		
Butter ...			Cowdung cakes		
Sweet oil			Electric charges		
Other oils used as food			Other fuel and lighting (e. g. spirit)		
Salt ...			Total expenditure on fuel and lighting		
Chillies ...			Actual rent (2) paid by the family including municipal tax		

NOTE—

(1) *Commodities* —Particulars should only be given for the commodities actually consumed by the family in the month referred to.

If any commodities were bought which are not entered in the list, please specify the names of such articles and give particulars of quantities and cost.

(2) *Rent.*—Where more than one family occupies the same tenement, special care should be taken to enter only the share of the rent actually paid by the family to which the budget refers.

Expenditure on clothing, shoes and other necessaries not bought every month

ARTICLES (*Note*:—The list of articles should be extended so as to include all those in use)	Number of articles in use	Cost per article when bought		Total cost of articles		Estimated number of months that each article will last	Estimated cost per month on average		
1	2	3		4		5	6		
CLOTHING	No.	Rs.	a.	Rs.	a.	Months	Rs.	a.	p.
FOR MEN									
Dhotis ..									
Shorts ...									
Lungis ...									
Pants ...									
Pagris or phetas ...									
Caps ...									
Bandis ...									
Pahirans ...									
Shirts ...									
Ganjifrocks (vests) ...									
Coats ...									
Handkerchiefs ...									
Socks ...									
Total ...									

FOR WOMEN								
Saris								
Odhanis								
Cholis								
Pyjamas								
Sadaras or kurtas								
Ghagaras								
Vijars								
Cloth for Blouses								
Total								
FOR MALE CHILDREN								
Dhotis								
Pants								
Half Shirts or Kurtas								
Coats								
Zablis								
Baby Caps								
Total								

Expenditure on clothing, shoes and other necessaries not bought every month—contd.

ARTICLES (*Note*:—The list of articles should be extended so as to include all those in use) 1	Number of articles in use 2	Cost per article when bought 3		Total cost of articles 4		Estimated number of months that each article will last 5	Estimated cost per month on average 6		
	No.	Rs.	a.	Rs.	a.	Months.	Rs.	a.	p.
CLOTHING—*contd.*									
FOR FEMALE CHILDREN									
Saris ...									
Odhanis ...									
Cholis and Polkas ...									
Ghagaras and Parkars ...									
Zablis ...									
Baby caps ...									
Total ...									
Total expenditure on all clothing ...									

UMBRELLAS								
SHOES OR SANDALS								
FOR MEN								
FOR WOMEN...								
FOR CHILDREN								
Total ...								
Total expenditure on Clothing, Umbrellas and Shoes or Sandals								

Instructions for filling up form on page 4 and 5

List of articles.—All the articles used should be entered in column 1. The articles named are only inserted as samples.

Number of articles.—In column 2 should be entered the number of each article actually in use at the time investigation, including changes of dress, possessed by the family.

Cost per article.—In column 3 should be entered the price per article when purchased. If similar articles were purchased at different prices, the prices should be entered separately and the number of articles at each price shown.

Total cost.—Column 4 will be obtained by multiplying the number in column 2 by the price in column 3.

Duration of articles.—In column 5 should be entered the total time which each article may be expected to last from the date of its purchase to the date when it is finally discarded as useless.

Cost per month.—Column 6 will be obtained by dividing the total cost in column 4 by the number entered in column 5.

Actual expenditure incurred on bedding and household necessaries during the month of enquiry

ARTICLES (The list of articles should be extended so as to include all those in use)	Number of articles purchased	Total cost of articles		
	No,	Rs.	as.	ps.
BEDDING*				
Cot or Charpoy ...				
Mattresses (Godadas and takias) ...				
Blankets ...				
Sheets ...				
Total ...				

HOUSEHOLD NECESSARIES*			
Cooking pots ...			
Earthen pots ...			
Lamps and Chimnies ...			
Pictures ...			
Furniture (chairs, table, swing) ...			
Other household necessaries			
Total ...			
Total monthly expenditure on bedding and household necessaries ...			

* Here show only the item purchased during the month of enquiry and the expenditure incurred on them. Please state whether brass, aluminium or enamelled iron utensils are more in use.

Miscellaneous Expenditure

Items	Cost per month			Items	Cost per month		
	Rs.	a.	p.		Rs.	a.	p.
Tooth-sticks (datan) ...				Trade union subscriptions ...			
Tooth powder ...				Other subscriptions ...			
Barber ...				Amusements (cinema, theatres, etc.) ...			
Dhobi ...				Medical fees and medicines ...			
Soap for washing ...				Education (books and fees) ...			
Toilet soap ...				Travelling expenses to and from work ...			
Hair oils ...				Snuff ...			
Pan-supari ...				Gifts and presents ...			
Chewing tobacco ...				Charity ...			
Bidis and matches ...				Servants ...			
Cigarettes ...				Land revenue ..			
Gadgu for hukka ...				Income-tax ...			
Tobacco ...				Lottery ...			
Liquor* ...				Luxuries such as watches, toys etc. ...			
Toddy ...				Flowers, etc. for worship ..			
Bhang ...				Travelling expenses to and from native place ...			
Opium ...				Payments to dependent members of family not residing at address given ...			
Bangles ...				Here note any other items of miscellaneous expenditure—			
Combs ...							

Postage ...							
Charges for grinding cereals and pulses ...							
Tinning ...							
Newspapers ...				Total monthly expenditure on miscellaneous items ...			

Summary of Income and Expenditure

	Rs.	a.	p.
Total Family Income in the month investigated (see page 2)			
Expenditure			
Food			
Fuel and lighting			
Rent			
Clothing, Umbrellas and shoes or Sandals			
Bedding and household necessaries			
Miscellaneous expenditure			
Total expenditure of family in the month ...			
Excess of income over expenditure			
How are the savings invested ... ... or			
How is the excess of expenditure over income met			

* State whether the liquor consumed is country or foreign.

Indebtedness

Causes. To what extent due to expenditure incurred on	Amount	From whom borrowed	Rate of interest per cent. per annum	Nature of security	Terms of repayment
	Rs. a. p.				
(*a*) Festival ...					
(*b*) Marriages ...					
(*c*) Funerals ...					
(*d*) Sickness ...					
(*e*) Unemployment ...					
(*f*) Other causes ...					
Total debt ...					

Housing—

(*i*) Landlord*

(*ii*) (*a*)

Number of rooms occupied by the family (including closed verandah)	Actual rent paid by the family (including taxes)		Approximate dimensions of each room (length, breadth and height)	Total persons in the tenement (including lodgers
	Rs.	a.		

(*b*) Dimensions of open verandah, if any … ______

(*iii*) Is the tenement shared by two or more families ?

If so give the budget numbers of such families ______

(*iv*) General sanitary condition ______

---

* State whether Government, Municipality, Employer, private Landlord or owned by self.

(2) An Inquiry into the Effects of the Inflation and Price Control on Middle Class Families in Bombay City.

It is well known that the middle classes and particularly those with fixed incomes have suffered the most because of high prices since the war. Various measures of control instituted by Government are introduced to bring relief to those in difficulties. Though it is claimed on behalf of the departments concerned that these controls have been successful, it is not possible to have a true picture of the situation unless a first hand investigation is made in a disinterested way into the family budgets of the people. It is possible to approach all classes of people for the necessary information. The object of this inquiry is to ascertain the effect of high prices and of controls on lower middle class families in Bombay City.

Lower middle class families are defined as those with incomes from Rs. 100 to 300 per month. It is proposed to select families from persons with fixed incomes as well as from the professions. At the same time efforts will be made to see that the families are drawn from different communities and different wards of the City. It is thus expected to get a representative sample on the basis of which conclusions can be drawn.

In view of the importance of the inquiry it is hoped that those who are approached for this information will be good enough to receive the investigators cordially and render them all assistance in their work. The investigators are post-graduate students in Economics in this School and are specially trained for this work. The investigation is being done by the Department of Economics of this School with the help of a grant from the University of Bombay.

It may be added that all information will be treated as confidential and that no name will be published.

SCHOOL OF ECONOMICS AND SOCIOLOGY,
UNIVERSITY OF BOMBAY,
Bombay 1-12-1944.

C. N. VAKIL
Director

Name

Address

Occupation of the head of the family :

(1) Did you live comfortably before the war i. e. could you make two ends meet easily before the war ? Could you save during the pre-war year ?

(2) How much has your income increased during the war ?

(3) Have you practised any economies on necessary articles during the war ? Please mention them in detail. How much has your expenditure increased during the war ?

(4) With these economies, have you been able to keep your expenditure within your income ? Or, had you to draw on past savings or/and incur fresh borrowing for meeting even this reduced expenditure ?

(5) Do you get any special facilities from your employers or any other organisations regarding your supplies and their prices? From whom and what sort of facilities do you get?

	Adults		Children below 14		Total	
	1939	1944	1939	1944	1939	1944
Members of the family living with the Heads:						
Members of the family and dependents living outside:						

Number of earners in the family.

1939

1944

No.	Basic pay		Dearness allowance, etc.	Subsidiary income with source		Total	
	1939	1944	1944	1939	1944	1939	1944
Total							

Commodities	1939			1944							Remarks
	Quantity (1)	Rate (2)	Expenses (3)	Total Quantity bought (4)	Quantity bought at controlled rates (5)	Controlled rates (6)	Expenditure on control purchases (7)	Other prices (8)	Other Expenses (9)	Total Expenses (10=7+9)	
FOOD											
Rice (Paylees)											
Wheat ,,											
Bajri ,,											
Jawar ,,											
Other cereals ,,											
Turdal ,,											
Other pulses ,,											
Bread (No.)											
Eggs (Dozens)											
Fish											
Beef (Lbs.)											
Mutton ,,											

53–54

Other meat	,,										
Milk	(Seers)										
Curd	,,										
Butter	,,										
Ghee	,,										
Vegetable Ghee	,,										
Mustard Oil	,,										
Cocoanut Oil	,,										
Gingily Oil	,,										
Other Oils used as food											
Potatoes	,,										
Vegetable (exclu-potatoes)											
Fruits (fresh & dried)											
Sugar	(Seers)										
Gur	,,										
Biscuits	,,										
Tea	(Lbs.)										
Coffee	,,										
Cocoa	,,										
Aerated Waters	(Bottles)										
Refreshments[1]											
Other foodstuffs[2]											
FOODSTUFFS	TOTAL										

---

[1] Refers to food taken in hotels.

[2] Sweetmeats, chocolates and other luxury foods are to be included here.

Commodities	1939			1944							Remarks
	Quantity (1)	Rate (2)	Expenses (3)	Total quantity bought (4)	Quantity bought at controlled rates (5)	Controlled rates (6)	Expenditure on controlled purchases (7)	Other prices (8)	Other Expenses (9)	Total Expenses (10+7=9)	
LIGHTING FUEL & RENT											
Coal (maunds)											
Charcoal ,,											
Firewood ,,											
Methylated spirit (bottles)											
Kerosene Oil (gallons)											
Electric charges											
Matches (boxes)											
* House rent (premium, if any)											
LIGHTING, FUEL & RENT TOTAL											

MISCELLANEOUS											
Tobacco (for chewing)											
Bidis (No.)											
Cigarettes & Cigars ,,											
Liquor, Spirits, etc.											
Toilet requisites											
Stationery											
Newspapers											
Medical fees											
Medicines											
Education[1]											
Amusements[2]											
Travel (Inside)											
Clubs											
Charity											
Gifts											
Wages to domestic servants											
Remittances to dependents											
Interest on debt											
Insurance (life)											
Provident fund											
Other extra expenditure											
Savings											

---

[1] Includes fees, expenditure in purchasing text books, etc.

[2] Expenditure on cinemas, concerts, cricket matches, games requisites should be included.

Commodities	1939			1944					Remarks
	Quantity bought	Rates	Total Expenditure	Quantity bought	Rates	Total expenditure	× of pur ehase in other market	Prices above controlled rate in other market	
ITEMS FOR WHICH NO MONTHLY EXPENDITURE IS INCURRED									
CLOTHES, FURNITURE, ETC.									
Cotton cloth (yards)									
Beds, bed sheets, etc.									
Mosquito curtains									
Woollen cloth (yards)									
Silk cloth ,,									
Artificial silk cloth ,,									

Overcoats (No.)								
Stiching charges								
Shoes (Pairs)								
Umbrellas (No.)								
Carpets, mats, chairs, etc.								
Cooking utensils								
Watches (No.)								
Fountain pens ,,								
Radios ,,								
Gramophones ,,								
Lamps (metal and glass) ,,								
Electric bulbs ,,								
Travel (outside)								
Holidays, festivals etc.								

Total expenditure:
1944

	Dec. 1939	Dec. 1943	Dec. 1944
Indebtedness (total)			
Savings (total)			
(a) Government loans			
(b) Post-office deposits			
(c) Other bank – savings			
(d) Life insurance			
(e) Industrial shares			
(f) Landed property			
(g) Gold and silver			
(h) Ornaments			
(i) Other			

NOTE: The figures for 1939 relate to August 1939 or nearest period.
The figures for 1944 relate to November 1944 or nearest period,

APPENDIX II

LOGARITHM AND THEIR USES

Calculations by Logarithms:—In statistical work very often tedious mathematical operations like multiplications, divisions, squaring, etc. are required to be done and if logarithms are made use of considerable saving of time and energy can be effected. It is likely that in the beginning the operation of logarithms might appear to be difficult but by practice it will be found to be very simple as well as useful.

A logarithm consists of the following two parts.—

(*a*) **The Characteristic**;—It is an integral part of a logarithm and is ascertained by inspection according to the following rule. The characteristic of a number greater than unity is positive and is one less than the number of digits to the left of the decimal point the characteristic of a number less than unity the negative and is numerically greater by one than the number of zeros following the decimal point. A bar is put over a negative characteristic.

The characteristic of

4723.4	is	3	·4724	is	$\bar{1}$
723.4	is	2	.0472	is	$\bar{2}$
23.4	is	1	.0047	is	$\bar{3}$
3.4	is	0	.0004	is	$\bar{4}$

(*b*) **The Mantissa**:—It is obtained by reference to a logarithmic table and can be taken with a number of digits but usually a four-figure fraction is found to be quite adequate for all practical purposes. It is always positive but the characteristic can be positive as well as negative.

By referring to a logarithmic table the mantissa will be ascertained and by inspection the characteristic will be put for the number as will be clear from the following illustration :—

The logarithm of	6145	is	3.7885
	614.50	,,	2.7885
	61.45	,,	1.7885

6.145	0.7885
0.6145	$\bar{1}.7885$
0.06145	$\bar{2}.7885$
0.006145	$\bar{3}.7885$

Computations by Logarithms:—The various logarithmic operations will be done as under :—

(*a*) **To Multiply Numbers** :—The multiplication of common numbers can be obtained by the addition of their logs.

$$\log (x \times y) = \log x + \log y$$

$$\therefore \; xy = \text{Antilog of } (\log x + \log y)$$

$$\begin{aligned} \log (52.57 \times 54.97) &= \log 52.57 + \log 54.97 \\ &= 1.7207 + 1.7401 \\ &= 3.4608 \end{aligned}$$

Its antilog is 2889.

∴ the product is 2889.

Multiply 52.57 by 0.05497

log .05497 = $\bar{2}.7401$ but for facilitating the adding or subtracting a logarithm with a negative characteristic we add and subtract 10 to make a positive characteristic, thus–

$$\begin{aligned} \log \; 52.57 &= 1.7207 \\ \log \; 0.05497 &= \underline{8.7401 - 10} \\ & 10.4608 - 10 \\ &= 0.4608 \end{aligned}$$

the antilog of 0.4608 is 2.889

∴ the product is 2.889

(*b*) **To Divide the Numbers**:—The division of common numbers can be obtained by substracting the logarithm of the divisor from the log of the dividend and the difference then shows the log of the quotient.

$$\log (x \div y) = \log x - \log y$$

$$\therefore \; x \div y = \text{Antilog} (\log x - \log y)$$

$$5257 \div 54.97$$

log 52.57=3.7207
log 54.97=1.7401
log of quotient=1.9806
Antilog of 9806=9.563
∴ quotient=95.63

(c) **To Raise a Number to Stated Power:**—A number can be raised to a stated power by multiplying the log of the number by the exponent of power. The product is the log of the repuired power.

$$\text{Log } x^n = n \log x$$
$$\therefore x^n = \text{Antilog}\,(n \log x)$$

Raise 2.5 to fourth power.
log 2.5=0,3997
multiplying by 4 we get 1.5916
Antilog of .5916=3.905
∴ $(2.5)^4 = 39.05$

(d) **To Extract the Root of a Number:**—The logarithm of the number should be divided by the index of the root and the quotient is the log of the required root.

$$\log \sqrt[n]{x} = \frac{1}{n} \log x$$
$$\sqrt[n]{x} = \text{Antilog}\left(\frac{\log x}{n}\right)$$

Find the 4th root of 39.05
log 39.05=1.5916
dividing by 4 we get 0.3979
Antilog of .3979=2.500
∴ the 4th root is 2.5

APPENDIX III

MATHEMATICAL TABLES

INSTRUCTIONS.

Table of Logarithms:—In this table mantissa of figures have been given to four digits only. Mantissa of a number is the same irrespective of the position of the decimal point. The characteristic is to be ascertained by inspection as explained already.

Table of Antilogarithms:—In this table are given the antilogarithms of the mantissa part of any given logarithm. By seeing the characteristic the position of the decimal point can be easily determined.

Table of Squares:—In this table two zeros have been omitted from 317 and one upto 316. If in the given figure the decimal point moves by one digit to the left the decimal point moves to the left by two digits in the square.

Table of Square-Roots:—In this table are given two square-roots for each number. The upper figure should be taken for an odd digit and lower one for an even digit. When the decimal point moves to the left by two digits in the given number the decimal point moves to the left by one digit in the square-root.

Table of Reciprocals:—If the decimal point moves by one digit to the right in the given number, the decimal point moves to the left in the reciprocal by one digit.

LOGARITHMS.

	0	1	2	3	4	5	6	7	8	9	1	2	3	4	5	6	7	8	9
10	0000	0043	0086	0128	0170						5	9	13	17	21	26	30	34	38
						0212	0253	0294	0334	0374	4	8	12	16	20	24	28	32	36
11	0414	0453	0492	0531	0569						4	8	12	16	20	23	27	31	35
						0607	0645	0682	0719	0755	4	7	11	15	18	22	26	29	33
12	0792	0828	0864	0899	0934						3	7	11	14	18	21	25	28	32
						0969	1004	1038	1072	1106	3	7	10	14	17	20	24	27	31
13	1139	1173	1206	1239	1271						3	6	10	13	16	19	23	26	29
						1303	1335	1367	1399	1430	3	7	10	13	16	19	22	25	29
14	1461	1492	1523	1553	1584						3	6	9	12	15	19	22	25	28
						1614	1644	1673	1703	1732	3	6	9	12	14	17	20	23	26
15	1761	1790	1818	1847	1875						3	6	9	11	14	17	20	23	26
						1903	1931	1959	1987	2014	3	6	8	11	14	17	19	22	25
16	2041	2068	2095	2122	2148						3	6	8	11	14	16	19	22	24
						2175	2201	2227	2253	2279	3	5	8	10	13	16	18	21	23
17	2304	2330	2355	2380	2405						3	5	8	10	13	15	18	20	23
						2430	2455	2480	2504	2529	3	5	8	10	12	15	17	20	22
18	2553	2577	2601	2625	2648						2	5	7	9	12	14	17	19	21
						2672	2695	2718	2742	2765	2	4	7	9	11	14	16	18	21
19	2788	2810	2833	2856	2878						2	4	7	9	11	13	16	18	20
						2900	2923	2945	2967	2989	2	4	6	8	11	13	15	17	19
20	3010	3032	3054	3075	3096	3118	3139	3160	3181	3201	2	4	6	8	11	13	15	17	19
21	3222	3243	3263	3284	3304	3324	3345	3365	3385	3404	2	4	6	8	10	12	14	16	18
22	3424	3444	3464	3483	3502	3522	3541	3560	3579	3598	2	4	6	8	10	12	14	15	17
23	3617	3636	3655	3674	3692	3711	3729	3747	3766	3784	2	4	6	7	9	11	13	15	17
24	3802	3820	3838	3856	3874	3892	3909	3927	3945	3962	2	4	5	7	9	11	12	14	16
25	3979	3997	4014	4031	4048	4065	4082	4099	4116	4133	2	3	5	7	9	10	12	14	15
26	4150	4166	4183	4200	4216	4232	4249	4265	4281	4298	2	3	5	7	8	10	11	13	15
27	4314	4330	4346	4362	4378	4393	4409	4425	4440	4456	2	3	5	6	8	9	11	13	14
28	4472	4487	4502	4518	4533	4548	4564	4579	4594	4609	2	3	5	6	8	9	11	12	14
29	4624	4639	4654	4669	4683	4698	4713	4728	4742	4757	1	3	4	6	7	9	10	12	13
30	4771	4786	4800	4814	4829	4843	4857	4871	4886	4900	1	3	4	6	7	9	10	11	13
31	4914	4928	4942	4955	4969	4983	4997	5011	5024	5038	1	3	4	6	7	8	10	11	12
32	5051	5065	5079	5092	5105	5119	5132	5145	5159	5172	1	3	4	5	7	8	9	11	12
33	5185	5198	5211	5224	5237	5250	5263	5276	5289	5302	1	3	4	5	6	8	9	10	12
34	5315	5328	5340	5353	5366	5378	5391	5403	5416	5428	1	3	4	5	6	8	9	10	11
35	5441	5453	5465	5478	5490	5502	5514	5527	5539	5551	1	2	4	5	6	7	9	10	11
36	5563	5575	5587	5599	5611	5623	5635	5647	5658	5670	1	2	4	5	6	7	8	10	11
37	5682	5694	5705	5717	5729	5740	5752	5763	5775	5786	1	2	3	5	6	7	8	9	10
38	5798	5809	5821	5832	5843	5855	5866	5877	5888	5899	1	2	3	5	6	7	8	9	10
39	5911	5922	5933	5944	5955	5966	5977	5988	5999	6010	1	2	3	4	5	7	8	9	10
40	6021	6031	6042	6053	6064	6075	6085	6096	6107	6117	1	2	3	4	5	6	8	9	10
41	6128	6138	6149	6160	6170	6180	6191	6201	6212	6222	1	2	3	4	5	6	7	8	9
42	6232	6243	6253	6263	6274	6284	6294	6304	6314	6325	1	2	3	4	5	6	7	8	9
43	6335	6345	6355	6365	6375	6385	6395	6405	6415	6425	1	2	3	4	5	6	7	8	9
44	6435	6444	6454	6464	6474	6484	6493	6503	6513	6522	1	2	3	4	5	6	7	8	9
45	6532	6542	6551	6561	6571	6580	6590	6599	6609	6618	1	2	3	4	5	6	7	8	9
46	6628	6637	6646	6656	6665	6675	6684	6693	6702	6712	1	2	3	4	5	6	7	7	8
47	6721	6730	6739	6749	6758	6767	6776	6785	6794	6803	1	2	3	4	5	5	6	7	8
48	6812	6821	6830	6839	6848	6857	6866	6875	6884	6893	1	2	3	4	4	5	6	7	8
49	6902	6911	6920	6928	6937	6946	6955	6964	6972	6981	1	2	3	4	4	5	6	7	8

LOGARITHMS.

	0	1	2	3	4	5	6	7	8	9	1	2	3	4	5	6	7	8	9
50	6990	6998	7007	7016	7024	7033	7042	7050	7059	7067	1	2	3	3	4	5	6	7	8
51	7076	7084	7093	7101	7110	7118	7126	7135	7143	7152	1	2	3	3	4	5	6	7	8
52	7160	7168	7177	7185	7193	7202	7210	7218	7226	7235	1	2	2	3	4	5	6	7	7
53	7243	7251	7259	7267	7275	7284	7292	7300	7308	7316	1	2	2	3	4	5	6	6	7
54	7324	7332	7340	7348	7356	7364	7372	7380	7388	7396	1	2	2	3	4	5	6	6	7
55	7404	7412	7419	7427	7435	7443	7451	7459	7466	7474	1	2	2	3	4	5	5	6	7
56	7482	7490	7497	7505	7513	7520	7528	7536	7543	7551	1	2	2	3	4	5	5	6	7
57	7559	7566	7574	7582	7589	7597	7604	7612	7619	7627	1	2	2	3	4	5	5	6	7
58	7634	7642	7649	7657	7664	7672	7679	7686	7694	7701	1	1	2	3	4	4	5	6	7
59	7709	7716	7723	7731	7738	7745	7752	7760	7767	7774	1	1	2	3	4	4	5	6	7
60	7782	7789	7796	7803	7810	7818	7825	7832	7839	7846	1	1	2	3	4	4	5	6	6
61	7853	7860	7868	7875	7882	7889	7896	7903	7910	7917	1	1	2	3	4	4	5	6	6
62	7924	7931	7938	7945	7952	7959	7966	7973	7980	7987	1	1	2	3	3	4	5	6	6
63	7993	8000	8007	8014	8021	8028	8035	8041	8048	8055	1	1	2	3	3	4	5	5	6
64	8062	8069	8075	8082	8089	8096	8102	8109	8116	8122	1	1	2	3	3	4	5	5	6
65	8129	8136	8142	8149	8156	8162	8169	8176	8182	8189	1	1	2	3	3	4	5	5	6
66	8195	8202	8209	8215	8222	8228	8235	8241	8248	8254	1	1	2	3	3	4	5	5	6
67	8261	8267	8274	8280	8287	8293	8299	8306	8312	8319	1	1	2	3	3	4	5	5	6
68	8325	8331	8338	8344	8351	8357	8363	8370	8376	8382	1	1	2	3	3	4	4	5	6
69	8388	8395	8401	8407	8414	8420	8426	8432	8439	8445	1	1	2	2	3	4	4	5	6
70	8451	8457	8463	8470	8476	8482	8488	8494	8500	8506	1	1	2	2	3	4	4	5	6
71	8513	8519	8525	8531	8537	8543	8549	8555	8561	8567	1	1	2	2	3	4	4	5	5
72	8573	8579	8585	8591	8597	8603	8609	8615	8621	8627	1	1	2	2	3	4	4	5	5
73	8633	8639	8645	8651	8657	8663	8669	8675	8681	8686	1	1	2	2	3	4	4	5	5
74	8692	8698	8704	8710	8716	8722	8727	8733	8739	8745	1	1	2	2	3	4	4	5	5
75	8751	8756	8762	8768	8774	8779	8785	8791	8797	8802	1	1	2	2	3	3	4	5	5
76	8808	8814	8820	8825	8831	8837	8842	8848	8854	8859	1	1	2	2	3	3	4	5	5
77	8865	8871	8876	8882	8887	8893	8899	8904	8910	8915	1	1	2	2	3	3	4	4	5
78	8921	8927	8932	8938	8943	8949	8954	8960	8965	8971	1	1	2	2	3	3	4	4	5
79	8976	8982	8987	8993	8998	9004	9009	9015	9020	9025	1	1	2	2	3	3	4	4	5
80	9031	9036	9042	9047	9053	9058	9063	9069	9074	9079	1	1	2	2	3	3	4	4	5
81	9085	9090	9096	9101	9106	9112	9117	9122	9128	9133	1	1	2	2	3	3	4	4	5
82	9138	9143	9149	9154	9159	9165	9170	9175	9180	9186	1	1	2	2	3	3	4	4	5
83	9191	9196	9201	9206	9212	9217	9222	9227	9232	9238	1	1	2	2	3	3	4	4	5
84	9243	9248	9253	9258	9263	9269	9274	9279	9284	9289	1	1	2	2	3	3	4	4	5
85	9294	9299	9304	9309	9315	9320	9325	9330	9335	9340	1	1	2	2	3	3	4	4	5
86	9345	9350	9355	9360	9365	9370	9375	9380	9385	9390	1	1	2	2	3	3	4	4	5
87	9395	9400	9405	9410	9415	9420	9425	9430	9435	9440	0	1	1	2	2	3	3	4	4
88	9445	9450	9455	9460	9465	9469	9474	9479	9484	9489	0	1	1	2	2	3	3	4	4
89	9494	9499	9504	9509	9513	9518	9523	9528	9533	9538	0	1	1	2	2	3	3	4	4
90	9542	9547	9552	9557	9562	9566	9571	9576	9581	9586	0	1	1	2	2	3	3	4	4
91	9590	9595	9600	9605	9609	9614	9619	9624	9628	9633	0	1	1	2	2	3	3	4	4
92	9638	9643	9647	9652	9657	9661	9666	9671	9675	9680	0	1	1	2	2	3	3	4	4
93	9685	9689	9694	9699	9703	9708	9713	9717	9722	9727	0	1	1	2	2	3	3	4	4
94	9731	9736	9741	9745	9750	9754	9759	9763	9768	9773	0	1	1	2	2	3	3	4	4
95	9777	9782	9786	9791	9795	9800	9805	9809	9814	9818	0	1	1	2	2	3	3	4	4
96	9823	9827	9832	9836	9841	9845	9850	9854	9859	9863	0	1	1	2	2	3	3	4	4
97	9868	9872	9877	9881	9886	9890	9894	9899	9903	9908	0	1	1	2	2	3	3	4	4
98	9912	9917	9921	9926	9930	9934	9939	9943	9948	9952	0	1	1	2	2	3	3	4	4
99	9956	9961	9965	9969	9974	9978	9983	9987	9991	9996	0	1	1	2	2	3	3	3	4

ANTILOGARITHMS.

	0	1	2	3	4	5	6	7	8	9	1	2	3	4	5	6	7	8	9
·00	1000	1002	1005	1007	1009	1012	1014	1016	1019	1021	0	0	1	1	1	1	2	2	2
·01	1023	1026	1028	1030	1033	1035	1038	1040	1042	1045	0	0	1	1	1	1	2	2	2
·02	1047	1050	1052	1054	1057	1059	1062	1064	1067	1069	0	0	1	1	1	1	2	2	2
·03	1072	1074	1076	1079	1081	1084	1086	1089	1091	1094	0	0	1	1	1	1	2	2	2
·04	1096	1099	1102	1104	1107	1109	1112	1114	1117	1119	0	1	1	1	1	2	2	2	2
·05	1122	1125	1127	1130	1132	1135	1138	1140	1143	1146	0	1	1	1	1	2	2	2	2
·06	1148	1151	1153	1156	1159	1161	1164	1167	1169	1172	0	1	1	1	1	2	2	2	2
·07	1175	1178	1180	1183	1186	1189	1191	1194	1197	1199	0	1	1	1	1	2	2	2	2
·08	1202	1205	1208	1211	1213	1216	1219	1222	1225	1227	0	1	1	1	1	2	2	2	3
·09	1230	1233	1236	1239	1242	1245	1247	1250	1253	1256	0	1	1	1	1	2	2	2	3
·10	1259	1262	1265	1268	1271	1274	1276	1279	1282	1285	0	1	1	1	1	2	2	2	3
·11	1288	1291	1294	1297	1300	1303	1306	1309	1312	1315	0	1	1	1	2	2	2	2	3
·12	1318	1321	1324	1327	1330	1334	1337	1340	1343	1346	0	1	1	1	2	2	2	2	3
·13	1349	1352	1355	1358	1361	1365	1368	1371	1374	1377	0	1	1	1	2	2	2	3	3
·14	1380	1384	1387	1390	1393	1396	1400	1403	1406	1409	0	1	1	1	2	2	2	3	3
·15	1413	1416	1419	1422	1426	1429	1432	1435	1439	1442	0	1	1	1	2	2	2	3	3
·16	1445	1449	1452	1455	1459	1462	1466	1469	1472	1476	0	1	1	1	2	2	2	3	3
·17	1479	1483	1486	1489	1493	1496	1500	1503	1507	1510	0	1	1	1	2	2	2	3	3
·18	1514	1517	1521	1524	1528	1531	1535	1538	1542	1545	0	1	1	1	2	2	2	3	3
·19	1549	1552	1556	1560	1563	1567	1570	1574	1578	1581	0	1	1	1	2	2	3	3	3
·20	1585	1589	1592	1596	1600	1603	1607	1611	1614	1618	0	1	1	1	2	2	3	3	3
·21	1622	1626	1629	1633	1637	1641	1644	1648	1652	1656	0	1	1	2	2	2	3	3	3
·22	1660	1663	1667	1671	1675	1679	1683	1687	1690	1694	0	1	1	2	2	2	3	3	3
·23	1698	1702	1706	1710	1714	1718	1722	1726	1730	1734	0	1	1	2	2	2	3	3	4
·24	1738	1742	1746	1750	1754	1758	1762	1766	1770	1774	0	1	1	2	2	2	3	3	4
·25	1778	1782	1786	1791	1795	1799	1803	1807	1811	1816	0	1	1	2	2	2	3	3	4
·26	1820	1824	1828	1832	1837	1841	1845	1849	1854	1858	0	1	1	2	2	3	3	3	4
·27	1862	1866	1871	1875	1879	1884	1888	1892	1897	1901	0	1	1	2	2	3	3	3	4
·28	1905	1910	1914	1919	1923	1928	1932	1936	1941	1945	0	1	1	2	2	3	3	4	4
·29	1950	1954	1959	1963	1968	1972	1977	1982	1986	1991	0	1	1	2	2	3	3	4	4
·30	1995	2000	2004	2009	2014	2018	2023	2028	2032	2037	0	1	1	2	2	3	3	4	4
·31	2042	2046	2051	2056	2061	2065	2070	2075	2080	2084	0	1	1	2	2	3	3	4	4
·32	2089	2094	2099	2104	2109	2113	2118	2123	2128	2133	0	1	1	2	2	3	3	4	4
·33	2138	2143	2148	2153	2158	2163	2168	2173	2178	2183	0	1	1	2	2	3	3	4	4
·34	2188	2193	2198	2203	2208	2213	2218	2223	2228	2234	1	1	2	2	3	3	4	4	5
·35	2239	2244	2249	2254	2259	2265	2270	2275	2280	2286	1	1	2	2	3	3	4	4	5
·36	2291	2296	2301	2307	2312	2317	2323	2328	2333	2339	1	1	2	2	3	3	4	4	5
·37	2344	2350	2355	2360	2366	2371	2377	2382	2388	2393	1	1	2	2	3	3	4	4	5
·38	2399	2404	2410	2415	2421	2427	2432	2438	2443	2449	1	1	2	2	3	3	4	4	5
·39	2455	2460	2466	2472	2477	2483	2489	2495	2500	2506	1	1	2	2	3	3	4	5	5
·40	2512	2518	2523	2529	2535	2541	2547	2553	2559	2564	1	1	2	2	3	4	4	5	5
·41	2570	2576	2582	2588	2594	2600	2606	2612	2618	2624	1	1	2	2	3	4	4	5	5
·42	2630	2636	2642	2649	2655	2661	2667	2673	2679	2685	1	1	2	2	3	4	4	5	6
·43	2692	2698	2704	2710	2716	2723	2729	2735	2742	2748	1	1	2	3	3	4	4	5	6
·44	2754	2761	2767	2773	2780	2786	2793	2799	2805	2812	1	1	2	3	3	4	4	5	6
·45	2818	2825	2831	2838	2844	2851	2858	2864	2871	2877	1	1	2	3	3	4	5	5	6
·46	2884	2891	2897	2904	2911	2917	2924	2931	2938	2944	1	1	2	3	3	4	5	5	6
·47	2951	2958	2965	2972	2979	2985	2992	2999	3006	3013	1	1	2	3	3	4	5	5	6
·48	3020	3027	3034	3041	3048	3055	3062	3069	3076	3083	1	1	2	3	4	4	5	6	6
·49	3090	3097	3105	3112	3119	3126	3133	3141	3148	3155	1	1	2	3	4	4	5	6	6

ANTILOGARITHMS.

	0	1	2	3	4	5	6	7	8	9	1	2	3	4	5	6	7	8	9
·50	3162	3170	3177	3184	3192	3199	3206	3214	3221	3228	1	1	2	3	4	4	5	6	7
·51	3236	3243	3251	3258	3266	3273	3281	3289	3296	3304	1	2	2	3	4	5	5	6	7
·52	3311	3319	3327	3334	3342	3350	3357	3365	3373	3381	1	2	2	3	4	5	5	6	7
·53	3388	3396	3404	3412	3420	3428	3436	3443	3451	3459	1	2	2	3	4	5	6	6	7
·54	3467	3475	3483	3491	3499	3508	3516	3524	3532	3540	1	2	2	3	4	5	6	6	7
·55	3548	3556	3565	3573	3581	3589	3597	3606	3614	3622	1	2	2	3	4	5	6	7	7
·56	3631	3639	3648	3656	3664	3673	3681	3690	3698	3707	1	2	3	3	4	5	6	7	8
·57	3715	3724	3733	3741	3750	3758	3767	3776	3784	3793	1	2	3	3	4	5	6	7	8
·58	3802	3811	3819	3828	3837	3846	3855	3864	3873	3882	1	2	3	4	4	5	6	7	8
·59	3890	3899	3908	3917	3926	3936	3945	3954	3963	3972	1	2	3	4	5	5	6	7	8
·60	3981	3990	3999	4009	4018	4027	4036	4046	4055	4064	1	2	3	4	5	6	6	7	8
·61	4074	4083	4093	4102	4111	4121	4130	4140	4150	4159	1	2	3	4	5	6	7	8	9
·62	4169	4178	4188	4198	4207	4217	4227	4236	4246	4256	1	2	3	4	5	6	7	8	9
·63	4266	4276	4285	4295	4305	4315	4325	4335	4345	4355	1	2	3	4	5	6	7	8	9
·64	4365	4375	4385	4395	4406	4416	4426	4436	4446	4457	1	2	3	4	5	6	7	8	9
·65	4467	4477	4487	4498	4508	4519	4529	4539	4550	4560	1	2	3	4	5	6	7	8	9
·66	4571	4581	4592	4603	4613	4624	4634	4645	4656	4667	1	2	3	4	5	6	7	9	10
·67	4677	4688	4699	4710	4721	4732	4742	4753	4764	4775	1	2	3	4	5	7	8	9	10
·68	4786	4797	4808	4819	4831	4842	4853	4864	4875	4887	1	2	3	4	6	7	8	9	10
·69	4898	4909	4920	4932	4943	4955	4966	4977	4989	5000	1	2	3	5	6	7	8	9	10
·70	5012	5023	5035	5047	5058	5070	5082	5093	5105	5117	1	2	4	5	6	7	8	9	11
·71	5129	5140	5152	5164	5176	5188	5200	5212	5224	5236	1	2	4	5	6	7	8	10	11
·72	5248	5260	5272	5284	5297	5309	5321	5333	5346	5358	1	2	4	5	6	7	9	10	11
·73	5370	5383	5395	5408	5420	5433	5445	5458	5470	5483	1	3	4	5	6	8	9	10	11
·74	5495	5508	5521	5534	5546	5559	5572	5585	5598	5610	1	3	4	5	6	8	9	10	12
·75	5623	5636	5649	5662	5675	5689	5702	5715	5728	5741	1	3	4	5	7	8	9	10	12
·76	5754	5768	5781	5794	5808	5821	5834	5848	5861	5875	1	3	4	5	7	8	9	11	12
·77	5888	5902	5916	5929	5943	5957	5970	5984	5998	6012	1	3	4	5	7	8	10	11	12
·78	6026	6039	6053	6067	6081	6095	6109	6124	6138	6152	1	3	4	6	7	8	10	11	13
·79	6166	6180	6194	6209	6223	6237	6252	6266	6281	6295	1	3	4	6	7	9	10	11	13
·80	6310	6324	6339	6353	6368	6383	6397	6412	6427	6442	1	3	4	6	7	9	10	12	13
·81	6457	6471	6486	6501	6516	6531	6546	6561	6577	6592	2	3	5	6	8	9	11	12	14
·82	6607	6622	6637	6653	6668	6683	6699	6714	6730	6745	2	3	5	6	8	9	11	12	14
·83	6761	6776	6792	6808	6823	6839	6855	6871	6887	6902	2	3	5	6	8	9	11	13	14
·84	6918	6934	6950	6966	6982	6998	7015	7031	7047	7063	2	3	5	6	8	10	11	13	15
·85	7079	7096	7112	7129	7145	7161	7178	7194	7211	7228	2	3	5	7	8	10	12	13	15
·86	7244	7261	7278	7295	7311	7328	7345	7362	7379	7396	2	3	5	7	8	10	12	13	15
·87	7413	7430	7447	7464	7482	7499	7516	7534	7551	7568	2	3	5	7	9	10	12	14	16
·88	7586	7603	7621	7638	7656	7674	7691	7709	7727	7745	2	4	5	7	9	11	12	14	16
·89	7762	7780	7798	7816	7834	7852	7870	7889	7907	7925	2	4	5	7	9	11	13	14	16
·90	7943	7962	7980	7998	8017	8035	8054	8072	8091	8110	2	4	6	7	9	11	13	15	17
·91	8128	8147	8166	8185	8204	8222	8241	8260	8279	8299	2	4	6	8	9	11	13	15	17
·92	8318	8337	8356	8375	8395	8414	8433	8453	8472	8492	2	4	6	8	10	12	14	15	17
·93	8511	8531	8551	8570	8590	8610	8630	8650	8670	8690	2	4	6	8	10	12	14	16	18
·94	8710	8730	8750	8770	8790	8810	8831	8851	8872	8892	2	4	6	8	10	12	14	16	18
·95	8913	8933	8954	8974	8995	9016	9036	9057	9078	9099	2	4	6	8	10	12	15	17	19
·96	9120	9141	9162	9183	9204	9226	9247	9268	9290	9311	2	4	6	8	11	13	15	17	19
·97	9333	9354	9376	9397	9419	9441	9462	9484	9506	9528	2	4	7	9	11	13	15	17	20
·98	9550	9572	9594	9616	9638	9661	9683	9705	9727	9750	2	4	7	9	11	13	16	18	20
·99	9772	9795	9817	9840	9863	9886	9908	9931	9954	9977	2	5	7	9	11	14	16	18	20

SQUARES

	0	1	2	3	4	5	6	7	8	9
10	1000	1020	1040	1061	1082	1103	1124	1145	1166	1188
11	1210	1232	1254	1277	1300	1323	1346	1369	1392	1416
12	1440	1464	1488	1513	1538	1563	1588	1613	1638	1664
13	1690	1716	1742	1769	1796	1823	1850	1877	1904	1932
14	1960	1988	2016	2045	2074	2103	2132	2161	2190	2220
15	2250	2280	2310	2341	2372	2403	2434	2465	2496	2528
16	2560	2592	2624	2657	2690	2723	2756	2789	2822	2856
17	2890	2924	2958	2993	3028	3063	3098	3133	3168	3204
18	3240	3276	3312	3349	3386	3423	3460	3497	3534	3572
19	3610	3648	3686	3725	3764	3803	3842	3881	3920	3960
20	4000	4040	4080	4121	4162	4203	4244	4285	4326	4368
21	4410	4452	4494	4537	4580	4623	4666	4709	4752	4796
22	4840	4884	4928	4973	5018	5063	5108	5153	5198	5244
23	5290	5336	5382	5429	5476	5523	5570	5617	5664	5712
24	5760	5808	5856	5905	5954	6003	6052	6101	6150	6200
25	6250	6300	6350	6401	6452	6503	6554	6605	6656	6708
26	6760	6812	6864	6917	6970	7023	7076	7129	7182	7236
27	7290	7344	7398	7453	7508	7563	7618	7673	7728	7784
28	7840	7896	7952	8009	8066	8123	8180	8237	8294	8352
29	8410	8468	8526	8585	8644	8703	8762	8821	8880	8940
30	9000	9060	9120	9181	9242	9303	9364	9425	9486	9548
31	9610	9672	9734	9797	9860	9923	9986	1005	1011	1018
32	1024	1030	1037	1043	1050	1056	1063	1069	1076	1082
33	1089	1096	1102	1109	1116	1122	1129	1136	1142	1149
34	1156	1163	1170	1176	1183	1190	1197	1204	1211	1218
35	1225	1232	1239	1246	1253	1260	1267	1274	1282	1289
36	1296	1303	1310	1318	1325	1332	1340	1347	1354	1362
37	1369	1376	1384	1391	1399	1406	1414	1421	1429	1436
38	1444	1452	1459	1467	1475	1482	1490	1498	1505	1513
39	1521	1529	1537	1544	1552	1560	1568	1576	1584	1592
40	1600	1608	1616	1624	1632	1640	1648	1656	1665	1673

The position of the decimal point must be determined according to the instructions given

SQUARES

	0	1	2	3	4	5	6	7	8	9
41	1681	1689	1697	1706	1714	1722	1731	1739	1747	1756
42	1764	1772	1781	1789	1798	1806	1815	1823	1832	1840
43	1849	1858	1866	1875	1884	1892	1901	1910	1918	1927
44	1936	1945	1954	1962	1971	1980	1989	1998	2007	2016
45	2025	2034	2043	2052	2061	2070	2079	2088	2098	2107
46	2116	2125	2134	2144	2153	2162	2172	2181	2190	2200
47	2209	2218	2228	2237	2247	2256	2266	2275	2285	2294
48	2304	2314	2323	2333	2343	2352	2362	2372	2381	2391
49	2401	2411	2421	2430	2440	2450	2460	2470	2480	2490
50	2500	2510	2520	2530	2540	2550	2560	2570	2581	2591
51	2601	2611	2621	2632	2642	2652	2663	2673	2683	2694
52	2704	2714	2725	2735	2746	2756	2767	2777	2788	2798
53	2809	2820	2830	2841	2852	2862	2873	2884	2894	2905
54	2916	2927	2938	2948	2959	2970	2981	2992	3003	3014
55	3025	3036	3047	3058	3069	3080	3091	3102	3114	3125
56	3136	3147	3158	3170	3181	3192	3204	3215	3226	3238
57	3249	3260	3272	3283	3295	3306	3318	3329	3341	3352
58	3364	3376	3387	3399	3411	3422	3434	3446	3457	3469
59	3481	3493	3505	3516	3528	3540	3552	3564	3576	3588
60	3600	3612	3624	3636	3648	3660	3672	3684	3697	3709
61	3721	3733	3745	3758	3770	3782	3795	3807	3819	3832
62	3844	3856	3869	3881	3894	3906	3919	3931	3944	3956
63	3969	3982	3994	4007	4020	4032	4045	4058	4070	4083
64	4096	4109	4122	4134	4147	4160	4173	4186	4199	4212
65	4225	4238	4251	4264	4277	4290	4303	4316	4330	4343
66	4356	4369	4382	4396	4409	4422	4436	4449	4462	4476
67	4489	4502	4516	4529	4543	4556	4570	4583	4597	4610
68	4624	4638	4651	4665	4679	4692	4706	4720	4733	4747
69	4761	4775	4789	4802	4816	4830	4844	4858	4872	4886
70	4900	4914	4928	4942	4956	4970	4984	4998	5013	5027
71	5041	5055	5069	5084	5098	5112	5127	5141	5155	5170
72	5184	5198	5213	5227	5242	5256	5271	5285	5300	5314

The position of the decimal point must be determined according to the instructions given.

SQUARES

	0	1	2	3	4	5	6	7	8	9
73	5329	5344	5358	5373	5388	5402	5417	5432	5446	5461
74	5476	5491	5506	5520	5535	5550	5565	5580	5595	5610
75	5625	5640	5655	5670	5685	5700	5715	5730	5746	5761
76	5776	5791	5806	5822	5837	5852	5868	5883	5898	5914
77	5929	5944	5960	5975	5991	6006	6022	6037	6053	6068
78	6084	6100	6115	6131	6147	6162	6178	6194	6209	6225
79	6241	6257	6273	6288	6304	6320	6336	6352	6368	6384
80	6400	6416	6432	6448	6464	6480	6496	6512	6529	6545
81	6561	6577	6593	6610	6626	6642	6659	6675	6691	6708
82	6724	6740	6757	6773	6790	6806	6823	6839	6856	6872
83	6889	6906	6922	6939	6956	6972	6989	7006	7022	7039
84	7056	7073	7090	7106	7123	7140	7157	7174	7191	7208
85	7225	7242	7259	7276	7293	7310	7327	7344	7362	7379
86	7396	7413	7430	7448	7465	7482	7500	7517	7534	7552
87	7569	7586	7604	7621	7639	7656	7674	7691	7709	7726
88	7744	7762	7779	7797	7815	7832	7850	7868	7885	7903
89	7921	7939	7957	7974	7992	8010	8028	8046	8064	8082
90	8100	8118	8136	8154	8172	8190	8208	8226	8245	8263
91	8281	8299	8317	8336	8354	8372	8391	8409	8427	8446
92	8464	8482	8501	8519	8538	8556	8575	8593	8612	8630
93	8649	8668	8686	8705	8724	8742	8761	8780	8798	8817
94	8836	8855	8874	8892	8911	8930	8949	8968	8987	9006
95	9025	9044	9063	9082	9101	9120	9139	9158	9178	9197
96	9216	9235	9254	9274	9293	9312	9332	9351	9370	9390
97	9409	9428	9448	9467	9487	9506	9526	9545	9565	9584
98	9604	9624	9643	9663	9683	9702	9722	9742	9761	9781
99	9801	9821	9841	9860	9880	9900	9920	9940	9960	9980

The position of the decimal point must be determined according to the instructions given.

SQUARE ROOTS

	0	1	2	3	4	5	6	7	8	9
10	1000	1005	1010	1015	1020	1025	1030	1034	1039	1044
	3162	3178	3194	3209	3225	3240	3256	3271	3286	3302
11	1049	1054	1058	1063	1068	1072	1077	1082	1086	1091
	3317	3332	3347	3362	3376	3391	3406	3421	3435	3450
12	1095	1100	1105	1109	1114	1118	1122	1127	1131	1136
	3464	3479	3493	3507	3521	3536	3550	3564	3578	3592
13	1140	1145	1149	1153	1158	1162	1166	1170	1175	1179
	3606	3619	3633	3647	3661	3674	3688	3701	3715	3728
14	1183	1187	1192	1196	1200	1204	1208	1212	1217	1221
	3742	3755	3768	3782	3795	3808	3821	3834	3847	3860
15	1225	1229	1233	1237	1241	1245	1249	1253	1257	1261
	3873	3886	3899	3912	3924	3937	3950	3962	3975	3987
16	1265	1269	1273	1277	1281	1285	1288	1292	1296	1300
	4000	4012	4025	4037	4050	4062	4074	4087	4099	4111
17	1304	1308	1311	1315	1319	1323	1327	1330	1334	1338
	4123	4135	4147	4159	4171	4183	4195	4207	4219	4231
18	1342	1345	1349	1353	1356	1360	1364	1367	1371	1375
	4243	4254	4266	4278	4290	4301	4313	4324	4336	4347
19	1378	1382	1386	1389	1393	1396	1400	1404	1407	1411
	4359	4370	4382	4393	4405	4416	4427	4438	4450	4461
20	1414	1418	1421	1425	1428	1432	1435	1439	1442	1446
	4472	4483	4494	4506	4517	4528	4539	4550	4561	4572
21	1449	1453	1456	1459	1463	1466	1470	1473	1476	1480
	4583	4593	4604	4615	4626	4637	4648	4658	4669	4680
22	1483	1487	1490	1493	1497	1500	1503	1507	1510	1513
	4690	4701	4712	4722	4733	4743	4754	4764	4775	4785
23	1517	1520	1523	1526	1530	1533	1536	1539	1543	1546
	4796	4806	4817	4827	4837	4848	4858	4868	4879	4889
24	1549	1552	1556	1559	1562	1565	1568	1572	1575	1578
	4899	4909	4919	4930	4940	4950	4960	4970	4980	4990

The first significant figure and the position of the decimal point must be determined in accordance with the instructions given.

SQUARE ROOTS

	0	1	2	3	4	5	6	7	8	9
25	1581 5000	1584 5010	1587 5020	1591 5030	1594 5040	1597 5050	1600 5060	1603 5070	1606 5079	1609 5089
26	1612 5099	1616 5109	1619 5119	1622 5128	1625 5138	1628 5148	1631 5158	1634 5167	1637 5177	1640 5187
27	1643 5196	1646 5206	1649 5215	1652 5225	1655 5235	1658 5244	1661 5254	1664 5263	1667 5273	1670 5282
28	1673 5292	1676 5301	1679 5310	1682 5320	1685 5329	1688 5339	1691 5348	1694 5357	1697 5367	1700 5376
29	1703 5385	1706 5394	1709 5404	1712 5413	1715 5422	1718 5431	1720 5441	1723 5450	1726 5459	1729 5468
30	1732 5477	1735 5486	1738 5495	1741 5505	1744 5514	1746 5523	1749 5532	1752 5541	1755 5550	1758 5559
31	1761 5568	1764 5577	1766 5586	1769 5595	1772 5604	1775 5612	1778 5621	1780 5630	1783 5639	1786 5648
32	1789 5657	1792 5666	1794 5675	1797 5683	1800 5692	1803 5701	1806 5710	1808 5718	1811 5727	1814 5736
33	1817 5745	1819 5753	1822 5762	1825 5771	1828 5779	1830 5788	1833 5797	1836 5805	1838 5814	1841 5822
34	1844 5831	1847 5840	1849 5848	1852 5857	1855 5865	1857 5874	1860 5882	1863 5891	1865 5899	1868 5908
35	1871 5916	1873 5925	1876 5933	1879 5941	1881 5950	1884 5958	1887 5967	1889 5975	1892 5983	1895 5992
36	1897 6000	1900 6008	1903 6017	1905 6025	1908 6033	1910 6042	1913 6050	1916 6058	1918 6066	1921 6075
37	1924 6083	1926 6091	1929 6099	1931 6107	1934 6116	1936 6124	1939 6132	1942 6140	1944 6148	1947 6156
38	1949 6164	1952 6173	1954 6181	1957 6189	1960 6197	1962 6205	1965 6213	1967 6221	1970 6229	1972 6237
39	1975 6245	1977 6253	1980 6261	1982 6269	1985 6277	1987 6285	1990 6293	1992 6301	1995 6309	1997 6317

The first significant figure and the position of the decimal point must be determined in accordance with the instructions given.

SQUARE ROOTS

	0	1	2	3	4	5	6	7	8	9
40	2000	2002	2005	2007	2010	2012	2015	2017	2020	2022
	6325	6332	6340	6348	6356	6364	6372	6380	6387	6395
41	2025	2027	2030	2032	2035	2037	2040	2042	2045	2047
	6403	6411	6419	6427	6434	6442	6450	6458	6465	6473
42	2049	2052	2054	2057	2056	2062	2064	2066	2069	2071
	6481	6488	6496	6504	6512	6519	6527	6535	6542	6550
43	2074	2076	2078	2081	2083	2086	2088	2090	2093	2095
	6557	6565	6573	6580	6588	6595	6603	6611	6618	6626
44	2098	2100	2102	2105	2107	2110	2112	2114	2117	2119
	6633	6641	6648	6656	6663	6671	6678	6686	6693	6701
45	2121	2124	2126	2128	2131	2133	2135	2138	2140	2142
	6708	6716	6723	6731	6738	6745	6753	6760	6768	6775
46	2145	2147	2149	2152	2154	2156	2159	2161	2163	2166
	6782	6790	6797	6804	6812	6819	6826	6834	6841	6848
47	2168	2170	2173	2175	2177	2179	2182	2184	2186	2189
	6856	6863	6870	6877	6885	6892	6899	6907	6914	6921
48	2191	2193	2195	2198	2200	2202	2205	2207	2209	2211
	6928	6935	6943	6950	6957	6964	6971	6979	6986	6993
49	2214	2216	2218	2220	2223	2225	2227	2229	2232	2234
	7000	7007	7014	7021	7029	7036	7043	7050	7057	7064
50	2236	2238	2241	2243	2245	2247	2249	2252	2254	2256
	7071	7078	7085	7092	7099	7106	7113	7120	7127	7134
51	2258	2261	2263	2265	2267	2269	2272	2274	2276	2278
	7141	7148	7155	7162	7169	7176	7183	7190	7197	7204
52	2280	2283	2285	2287	2289	2291	2293	2296	2298	2300
	7211	7218	7225	7232	7239	7246	7253	7259	7266	7273
53	2302	2304	2307	2309	2311	2313	2315	2317	2319	2322
	7280	7287	7294	7301	7308	7314	7321	7328	7335	7342
54	2324	2326	2328	2330	2332	2335	2337	2339	2341	2343
	7348	7355	7362	7369	7376	7382	7389	7396	7403	7409

The first significant figure and the position of the decimal point must be determined in accordance with the instructions given.

SQUARE ROOTS

	0	1	2	3	4	5	6	7	8	9
55	2345	2347	2349	2352	2354	2356	2358	2360	2362	2364
	7416	7423	7430	7436	7443	7450	7457	7463	7470	7477
56	2366	2369	2371	2373	2375	2377	2379	2381	2383	2385
	7483	7490	7497	7503	7510	7517	7523	7530	7537	7543
57	2387	2390	2392	2394	2396	2398	2400	2402	2404	2406
	7550	7556	7563	7570	7576	7583	7589	7596	7603	7609
58	2408	2410	2412	2415	2417	2419	2421	2423	2425	2427
	7616	7622	7629	7635	7642	7649	7655	7662	7668	7675
59	2429	2431	2433	2435	2437	2439	2441	2443	2445	2447
	7681	7688	7694	7701	7707	7714	7720	7727	7733	7740
60	2449	2452	2454	2456	2458	2460	2462	2464	2466	2468
	7746	7752	7759	7765	7772	7778	7785	7791	7797	7804
61	2470	2472	2474	2476	2478	2480	2482	2484	2486	2488
	7810	7817	7823	7829	7836	7842	7849	7855	7861	7868
62	2490	2492	2494	2496	2498	2500	2502	2504	2506	2508
	7874	7880	7887	7893	7899	7906	7912	7918	7925	7931
63	2510	2512	2514	2516	2518	2520	2522	2524	2526	2528
	7937	7944	7950	7956	7962	7969	7975	7981	7987	7994
64	2530	2532	2534	2536	2538	2540	2542	2544	2546	2548
	8000	8006	8012	8019	8025	8031	8037	8044	8050	8056
65	2550	2551	2553	2555	2557	2559	2561	2563	2565	2567
	8062	8068	8075	8081	8087	8093	8099	8106	8112	8118
66	2569	2571	2573	2575	2577	2579	2581	2583	2585	2587
	8124	8130	8136	8142	8149	8155	8161	8167	8173	8179
67	2588	2590	2592	2594	2596	2598	2600	2602	2604	2606
	8185	8191	8198	8204	8210	8216	8222	8228	8234	8240
68	2608	2610	2612	2613	2615	2617	2619	2621	2623	2625
	8246	8252	8258	8264	8270	8276	8283	8289	8295	8301
69	2627	2629	2631	2632	2634	2636	2638	2640	2642	2644
	8307	8313	8319	8325	8331	8337	8343	8349	8355	8361

The first significant figure and the position of the decimal point be determined in accordance with the instructions given.

SQUARE ROOTS

	0	1	2	3	4	5	6	7	8	9
70	2646	2648	2650	2651	2653	2655	2657	2659	2661	2663
	8367	8373	8379	8385	8390	8396	8402	8408	8414	8420
71	2665	2666	2668	2670	2672	2674	2676	2678	2680	2681
	8426	8432	8438	8444	8450	8456	8462	8468	8473	8479
72	2683	2685	2687	2689	2691	2693	2694	2696	2698	2700
	8485	8491	8497	8503	8509	8515	8521	8526	8532	8538
73	2702	2704	2706	2707	2709	2711	2713	2715	2717	2718
	8544	8550	8556	8562	8567	8573	8579	8585	8591	8597
74	2720	2722	2724	2726	2728	2729	2731	2733	2735	2737
	8602	8608	8614	8620	8626	8631	8637	8643	8649	8654
75	2739	2740	2742	2744	2746	2748	2750	2751	2753	2755
	8660	8666	8672	8678	8683	8689	8695	8701	8706	8712
76	2757	2759	2760	2762	2764	2766	2768	2769	2771	2773
	8718	8724	8729	8735	8741	8746	8752	8758	8764	8769
77	2775	2777	2778	2780	2782	2784	2786	2787	2789	2791
	8775	8781	8786	8792	8798	8803	8809	8815	8820	8826
78	2793	2795	2796	2798	2800	2802	2804	2805	2807	2809
	8832	8837	8843	8849	8854	8860	8866	8871	8877	8883
79	2811	2812	2814	2816	2818	2820	2821	2823	2825	2827
	8888	8894	8899	8905	8911	8916	8922	8927	8933	8939
80	2828	2830	2832	2834	2835	2837	2839	2841	2843	2844
	8944	8950	8955	8961	8967	8972	8978	8983	8989	8994
81	2846	2848	2850	2851	2853	2855	2857	2858	2860	2862
	9000	9006	9011	9017	9022	9028	9033	9039	9044	9050
82	2864	2865	2867	2869	2871	2872	2874	2876	2877	2879
	9055	9061	9066	9072	9077	9083	9088	9094	9099	9105
83	2881	2883	2884	2886	2888	2890	2891	2893	2895	2897
	9110	9116	9121	9127	9132	9138	9143	9149	9154	9160
84	2898	2900	2902	2903	2905	2907	2909	2910	2912	2914
	9165	9171	9176	9182	9187	9192	9198	9203	9209	9214

The first significant figure and the position of the decimal point must be determined in accordance with the instructions given.

SQUARE ROOTS

	0	1	2	3	4	5	6	7	8	9
85	2915 9220	2917 9225	2919 9230	2921 9236	2922 9241	2924 9247	2926 9252	2927 9257	2929 9263	2931 9268
86	2933 9274	2934 9279	2936 9284	2938 9290	2939 9295	2941 9301	2943 9306	2944 9311	2946 9317	2948 9322
87	2950 9327	2951 9333	2953 9338	2955 9343	2956 9349	2958 9354	2960 9359	2961 9365	2963 9370	2965 9375
88	2966 9381	2968 9386	2970 9391	2972 9397	2973 9402	2975 9407	2977 9413	2978 9418	2980 9423	2982 9429
89	2983 9434	2985 9439	2987 9445	2988 9450	2990 9455	2992 9460	2993 9466	2995 9471	2997 9476	2998 9482
90	3000 9487	3002 9492	3003 9497	3005 9503	3007 9508	3008 9513	3010 9518	3012 9524	3013 9529	3015 9534
91	3017 9539	3018 9545	3020 9550	3022 9555	3023 9560	3025 9566	3027 9571	3028 9576	3030 9581	3032 9586
92	3033 9592	3035 9597	3036 9602	3038 9607	3040 9612	3041 9618	3043 9623	3045 9628	3046 9633	3048 9638
93	3050 9644	3051 9649	3053 9654	3055 9659	3056 9664	3058 9670	3059 9675	3061 9680	3063 9685	3064 9690
94	3066 9695	3068 9701	3069 9706	3071 9711	3072 9716	3074 9721	3076 9726	3077 9731	3079 9737	3081 9742
95	3082 9747	3084 9752	3085 9757	3087 9762	3089 9767	3090 9772	3092 9778	3094 9783	3095 9788	3097 9793
96	3098 9798	3100 9803	3102 9808	3103 9813	3105 9818	3106 9823	3108 9829	3110 9834	3111 9839	3113 9844
97	3114 9849	3116 9854	3118 9859	3119 9864	3121 9869	3122 9874	3124 9879	3126 9884	3127 9889	3129 9894
98	3130 9899	3132 9905	3134 9910	3135 9915	3137 9920	3138 9925	3140 9930	3142 9935	3143 9940	3145 9945
99	3146 9950	3148 9955	3150 9960	3151 9965	3153 9970	3154 9975	3156 9980	3158 9985	3159 9990	3161 9995

The first significant figure and the position of the decimal point be determined in accordance with the instructions given.

RECIPROCALS

	0	1	2	3	4	5	6	7	8	9
4·2	·2381	·2375	·2370	·2364	·2358	·2353	·2347	·2342	·2336	·2331
4·3	·2326	·2320	·2315	·2309	·2304	·2299	·2294	·2288	·2283	·2278
4·4	·2273	·2268	·2262	·2257	·2252	·2247	·2242	·2237	·2232	·2227
4·5	·2222	·2217	·2212	·2208	·2203	·2198	·2193	·2188	·2183	·2179
4·6	·2174	·2169	·2165	·2160	·2155	·2151	·2146	·2141	·2137	·2132
4·7	·2128	·2123	·2119	·2114	·2110	·2105	·2101	·2096	·2092	·2088
4·8	·2083	·2079	·2075	·2070	·2066	·2062	·2058	·2053	·2049	·2045
4·9	·2041	·2037	·2033	·2028	·2024	·2020	·2016	·2012	·2008	·2004
5·0	·2000	·1996	·1992	·1988	·1984	·1980	·1976	·1972	·1969	·1965
5·1	·1961	·1957	·1953	·1949	·1946	·1942	·1938	·1934	·1931	·1927
5·2	·1923	·1919	·1916	·1912	·1908	·1905	·1901	·1898	·1894	·1890
5·3	·1887	·1883	·1880	·1876	·1873	·1869	·1866	·1862	·1859	·1855
5·4	·1852	·1848	·1845	·1842	·1838	·1835	·1832	·1828	·1825	·1821
5·5	·1818	·1815	·1812	·1808	·1805	·1802	·1799	·1795	·1792	·1789
5·6	·1786	·1783	·1779	·1776	·1773	·1770	·1767	·1764	·1761	·1757
5·7	·1754	·1751	·1748	·1745	·1742	·1739	·1736	·1733	·1730	·1727
5·8	·1724	·1721	·1718	·1715	·1712	·1709	·1706	·1704	·1701	·1698
5·9	·1695	·1692	·1689	·1686	·1684	·1681	·1678	·1675	·1672	·1669
6·0	·1667	·1664	·1661	·1658	·1656	·1653	·1650	·1647	·1645	·1642
6·1	·1639	·1637	·1634	·1631	·1629	·1626	·1623	·1621	·1618	·1616
6·2	·1613	·1610	·1608	·1605	·1603	·1600	·1597	·1595	·1592	·1590
6·3	·1587	·1585	·1582	·1580	·1577	·1575	·1572	·1570	·1567	·1565
6·4	·1563	·1560	·1558	·1555	·1553	·1550	·1548	·1546	·1543	·1541
6·5	·1538	·1536	·1534	·1531	·1529	·1527	·1524	·1522	·1520	·1517
6·6	·1515	·1513	·1511	·1508	·1506	·1504	·1502	·1499	·1497	·1495
6·7	·1493	·1490	·1488	·1486	·1484	·1481	·1479	·1477	·1475	·1473
6·8	·1471	·1468	·1466	·1464	·1462	·1460	·1458	·1456	·1453	·1451
6·9	·1449	·1447	·1445	·1443	·1441	·1439	·1437	·1435	·1433	·1431
7·0	·1429	·1427	·1425	·1422	·1420	·1418	·1416	·1414	·1412	·1410
7·1	·1408	·1406	·1404	·1403	·1401	·1399	·1397	·1395	·1393	·1391
7·2	·1389	·1387	·1385	·1383	·1381	·1379	·1377	·1376	·1374	·1372
7·3	·1370	·1368	·1366	·1364	·1362	·1361	·1359	·1357	·1355	·1353
7·4	·1351	·1350	·1348	·1346	·1344	·1342	·1340	·1339	·1337	·1335

RECIPROCALS

	0	1	2	3	4	5	6	7	8	9
1·0	1·0000	·9901	·9804	·9709	·9615	·9524	·9434	·9346	·9259	·9174
1·1	·9091	·9009	·8929	·8850	·8772	·8696	·8621	·8547	·8475	·8403
1·2	·8333	·8264	·8197	·8130	·8065	·8000	·7937	·7874	·7813	·7752
1·3	·7692	·7634	·7576	·7519	·7463	·7407	·7353	·7299	·7246	·7194
1·4	·7143	·7092	·7042	·6993	·6944	·6897	·6849	·6803	·6757	·6711
1·5	·6667	·6623	·6579	·6536	·6494	·6452	·6410	·6369	·6329	·6289
1·6	·6250	·6211	·6173	·6135	·6098	·6061	·6024	·5988	·5952	·5917
1·7	·5882	·5848	·5814	·5780	·5747	·5714	·5682	·5650	·5618	·5587
1·8	·5556	·5525	·5495	·5464	·5435	·5405	·5376	·5348	·5319	·5291
1·9	·5263	·5236	·5208	·5181	·5155	·5128	·5102	·5076	·5051	·5025
2·0	·5000	·4975	·4950	·4926	·4902	·4878	·4854	·4831	·4808	·4785
2·1	·4762	·4739	·4717	·4695	·4673	·4651	·4630	·4608	·4587	·4566
2·2	·4545	·4525	·4505	·4484	·4464	·4444	·4425	·4405	·4386	·4367
2·3	·4348	·4329	·4310	·4292	·4274	·4255	·4237	·4219	·4202	·4184
2·4	·4167	·4149	·4132	·4115	·4098	·4082	·4065	·4049	·4032	·4016
2·5	·4000	·3984	·3968	·3953	·3937	·3922	·3906	·3891	·3876	·3861
2·6	·3846	·3831	·3817	·3802	·3788	·3774	·3759	·3745	·3731	·3717
2·7	·3704	·3690	·3676	·3663	·3650	·3636	·3623	·3610	·3597	·3584
2·8	·3571	·3559	·3546	·3534	·3521	·3509	·3497	·3484	·3472	·3460
2·9	·3448	·3436	·3425	·3413	·3401	·3390	·3378	·3367	·3356	·3344
3·0	·3333	·3322	·3311	·3300	·3289	·3279	·3268	·3257	·3247	·3236
3·1	·3226	·3215	·3205	·3195	·3185	·3175	·3165	·3155	·3145	·3135
3·2	·3125	·3115	·3106	·3096	·3086	·3077	·3067	·3058	·3049	·3040
3·3	·3030	·3021	·3012	·3003	·2994	·2985	·2976	·2967	·2959	·2950
3·4	·2941	·2933	·2924	·2915	·2907	·2899	·2890	·2882	·2874	·2865
3·5	·2857	·2849	·2841	·2833	·2825	·2817	·2809	·2801	·2793	·2786
3·6	·2778	·2770	·2762	·2755	·2747	·2740	·2732	·2725	·2717	·2710
3·7	·2703	·2695	·2688	·2681	·2674	·2667	·2660	·2653	·2646	·2639
3·8	·2632	·2625	·2618	·2611	·2604	·2597	·2591	·2584	·2577	·2571
3·9	·2564	·2558	·2551	·2545	·2538	·2532	·2525	·2519	·2513	·2506
4·0	·2500	·2494	·2488	·2481	·2475	·2469	·2463	·2457	·2451	·2445
4·1	·2439	·2433	·2427	·2421	·2415	·2410	·2404	·2398	·2392	·2387

RECIPROCALS

	0	1	2	3	4	5	6	7	8	9
7·5	·1333	·1332	·1330	·1328	·1326	·1325	·1323	·1321	·1319	·1318
7·6	·1316	·1314	·1312	·1311	·1309	·1307	·1305	·1304	·1302	·1300
7·7	·1299	·1297	·1295	·1294	·1292	·1290	·1289	·1287	·1285	·1284
7·8	·1282	·1280	·1279	·1277	·1276	·1274	·1272	·1271	·1269	·1267
7·9	·1266	·1264	·1263	·1261	·1259	·1258	·1256	·1255	·1253	·1252
8·0	·1250	·1248	·1247	·1245	·1244	·1242	·1241	·1239	·1238	·1236
8·1	·1235	·1233	·1232	·1230	·1229	·1227	·1225	·1224	·1222	·1221
8·2	·1220	·1218	·1217	·1215	·1214	·1212	·1211	·1209	·1208	·1206
8·3	·1205	·1203	·1202	·1200	·1199	·1198	·1196	·1195	·1193	·1192
8·4	·1190	·1189	·1188	·1186	·1185	·1183	·1182	·1181	·1179	·1178
8·5	·1176	·1175	·1174	·1172	·1171	·1170	·1168	·1167	·1166	·1164
8·6	·1163	·1161	·1160	·1159	·1157	·1156	·1155	·1153	·1152	·1151
8·7	·1149	·1148	·1147	·1145	·1144	·1143	·1142	·1140	·1139	·1138
8·8	·1136	·1135	·1134	·1133	·1131	·1130	·1129	·1127	·1126	·1125
8·9	·1124	·1122	·1121	·1120	·1119	·1117	·1116	·1115	·1114	·1112
9·0	·1111	·1110	·1109	·1107	·1106	·1105	·1104	·1103	·1101	·1100
9·1	·1099	·1098	·1096	·1095	·1094	·1093	·1092	·1091	·1089	·1088
9·2	·1087	·1086	·1085	·1083	·1082	·1081	·1080	·1079	·1078	·1076
9·3	·1075	·1074	·1073	·1072	·1071	·1070	·1068	·1067	·1066	·1065
9·4	·1064	·1063	·1062	·1060	·1059	·1058	·1057	·1056	·1055	·1054
9·5	·1053	·1052	·1050	·1049	·1048	·1047	·1046	·1045	·1044	·1043
9·6	·1042	·1041	·1040	·1038	·1037	·1036	·1035	·1034	·1033	·1032
9·7	·1031	·1030	·1029	·1028	·1027	·1026	·1025	·1024	·1022	·1021
9·8	·1020	·1019	·1018	·1017	·1016	·1015	·1014	·1013	·1012	·1011
9·9	·1010	·1009	·1008	·1007	·1006	·1005	·1004	·1003	·1002	·1001

APPENDIX IV

IMPORTANT FORMULAS AND SYMBOLS

$$\text{Mean} = \text{Arb. Mean} + \frac{\Sigma fd}{n} \times i$$

Mean of the combined series $= \dfrac{M_1 n_1 + M_2 n_2}{n_1 + n_2}$ where M_1 and M_2 are the arithmetic means of the two series and n_1 and n_2 totals of their frequencies.

$$\text{First or Lower Quartile} = l + \frac{\frac{(n+1)}{4} - cf}{fr} \times i$$

$$\text{Median or Second Quartile} = l + \frac{\frac{2(n+1)}{4} - cf}{fr} \times i$$

$$\text{Third or Upper Quartile} = l + \frac{\frac{3(n+1)}{4} - cf}{fr} \times i$$

Mode = Mean − 3 (Mean − Median), or

$$\text{Mode} = l + \frac{f_2}{f_2 + f_1} \times i$$

$$\text{Quartile Deviation} = \frac{Q_3 - Q_1}{2}$$

$$\text{Coefficient of Quartile Deviation} = \frac{Q_3 - Q_1}{Q_3 + Q_1}$$

$$\text{Mean Deviation} = \frac{\Sigma fd + c(n_1 - n_2) - q}{n} \times i$$

$$\text{Standard Deviation} = \sqrt{\frac{\Sigma fd^2}{n} - c^2} \times i$$

Coefficient of	Mean Deviation	Standard Deviation
based on Mean	$\frac{\text{M. D.}}{\text{Mean.}}$	$\frac{\text{S. D.}}{\text{Mean}}$

based on Median	$\frac{\text{M. D.}}{\text{Median}}$	$\frac{\text{S. D.}}{\text{Median}}$
based on Mode	$\frac{\text{M. D.}}{\text{Mode}}$	$\frac{\text{S. D.}}{\text{Mode}}$

$$\text{Skewness} = \frac{\text{Mean} - \text{Mode}}{\text{Standard Deviation}} \text{ or}$$

$$= \frac{3\,(\text{Mean} - \text{Median})}{\text{Standard Deviation.}}$$

$$\text{Coefficient of skewness} = \frac{Q_3 + Q_1 - 2\ \text{Median}}{\frac{Q_3 - Q_1}{2}}$$

$$\text{Coefficient of Correlation} = \frac{\Sigma xy}{n \times \text{S. D.}_x \times \text{S. D.}_y} \text{ or}$$

$$= \frac{\Sigma xy}{\sqrt{\Sigma x_2 \, x \, \Sigma y_2}}$$

$$\text{Coefficient of correlation (of a group)} = \frac{\frac{\Sigma xy}{n} - cx \times cy}{\text{S. D.}_x \times \text{S.D.}_y}$$

$$\text{Probable Error} = \frac{0{\cdot}6745\,(1 - r_2)}{\sqrt{n}}$$

$$\text{Modulus} = \text{S.D.} \times \sqrt{2}$$

$$\text{Coefficient of Variation} = \frac{\text{S. D.}}{\text{Mean}} \times 100$$

$$\text{Three Moments} = \frac{\Sigma fd}{n},\ \frac{\Sigma fd^2}{n},\ \frac{\Sigma \times d^3}{n}.$$

A.M. = Arithmetic Mean.

Ar. M. = Arbitrary Mean.

c (small) = Mean difference between the mean and the arbitrary mean i.e. $\frac{\Sigma fd}{n}$

C_f = Cumulative frequency of the class lower to one in which the quartile is situated.

d = Deviation from the Mean.

e = Absolute error

f = Frequency

f_1 = Number of items in the class above that containing the mode.

f_2 = Number of items in the class above that containing the mode.

f_r = Frequency of the class in which the quartile is situated.

i = Class interval.

j = Coefficient of skewness.

l = Lower limit of the class in which the quartile is situated.

M. D. = Mean Deviation.

M_e = Median.

M = Geometric Mean.

M_h = Harmonic Mean.

M_o = Mode.

n = Total number of frequencies.

n_1 = Number of items below the mean

n_2 = Number of items above the mean.

P. E. = Probable Error.

Q = Double the amount of 'fd' items falling between the classes covered by the mean and arbitrary mean (without taking signs into account).

Q. D = Quartile Deviation.

Q_1 = First or Lower quartile.

Q_2 = Second quartile or Median.

Q_3 = Third or Upper quartile.

r = Coefficient of correlation.

S. D. = Standard Deviation.

v = Coefficient of Variation.

v_1 = First moment.

v_3 = Second moment.

v_2 = Third moment.

w = Weights used.

x = Usually used to denote a variable.

y = Usually used to denote a variable.

xy = Multiplication of respective deviations of the two series from their mean or arbitrary mean as the case may be.

Z = **Mode.**

6 = Standard Deviation.

Σ = Total.

Σfd = Sum of frequencies multiplied by their respective deviations from the arbitrary mean.

Appendix V

List of some of the important publications containing statistics.

(A) *Indian*

Weekly (Government).

1. Indian Trade Journal.
2. Gazette of India.
3. Weekly Market Reports.

Monthly (Government).

1. Accounts relating to Sea-borne Trade and Navigation of India.
2. Cotton Spinning and Weaving in Indian Mills.
3. Monthly Statement of Wholesale Prices of certain selected articles in India.
4. Accounts relating to the Inland (Rail or River-borne) Trade of India.
5. Labour Gazette.
6. Monthly Statistics of Production of certain selected Industries of India.
7. Survey of Business Conditions issued by the Economic Adviser.
8. Bank-Statistical Summary.
9. Raw Cotton Trade Statistics.

Annual (Government).

1. Agricultural Statistics of India. Vol. I & II
2. Estimates of Area and Yield of principal crops in India.
3. Review of Trade of India.
4. Statistics relating to Banks in India.
5. Statistical Tables relating to Joint Stock Companies.
6. Statistical Statement relating to the Co-operative Movement in India.
7. Report on the production of Tea, Rubber, Coal, Coffee.
8. Budgets of India and Provincial Governments.
9. Moral and Material Progress Report of the Government of India.
10. Administration Report of Provincial Governments.
11. Annual Administration Report of Railways in India.
12. Report of the Irrigation Departments of the various provinces.
13. Annual Reports of Post and Telegraph Department.
14. Reports of the Forest Department of the Various Provinces.
15. Indian Weather Review.
16. Agricultural and Animal Husbandry in India (Imperial Council of Agriculture Research).
17. Annual Report of the Chief Inspector of Mines for India.
18. Annual Statement of Mineral Production in India.
19. Statistical Abstract for India.
20. Agricultural and Animal Husbandry in India.
21. Annual Report on the Department of Fisheries.
22. Statistics of Factories.
23. Annual Statement of Sea-borne Trade.

Periodical Publications (Government).

1. Large Industrial Establishments in India. (Biennial).
2. Statistical Abstract of India (Triennial).
3. Crop Forecasts of Wheat, Rice, Cotton, Linseed, Groundnut, Sugar-cane (as supplement to Indian Trade Journal.
4. Census Reports (Decennial).
5. Settlement Reports.
6. District Gazetteers.
7. Working class Family Budgets.
8. Varietal Cotton Map of India.
9. Crop Atlas of India.
10. Guide to Current Official Statistics.
11. Report of the Average Yield per acre of provincial crops in India (Quinquennial).
12. Live Stock statistics (Quinquennial).
13. Marketing Reports.
14. Triennial Review of Irrigation.
15. Index Number of prices of Exported and Imported Articles.
16. The Indian Insurances Year Book.
17. Index Numbers of Indian Prices (Quenquennial).

Reports of important committees and commissions of statistical importance.

1. Datta's Report on the Rice of Prices in India. (1912).
2. Report of the Economic Enquiry Committee (1925).
3. Report of the Bowley Robertson Committee.
4. Report of the Royal Commission on Indian Agriculture.
5. Report of the Taxation Inquiry Committee.
6. Reports of the Commissions on Indian Currency and Exchange.
7. Report of the Central Banking Inquiry Committee.
8. Report of the Royal Commission on Indian Labour.
9. Industrial Commission Report.
10. Tariff Board Reports.

Other Publications

1. Capital.
2. Commerce.
3. Indian Finance
4. Indian Year Book.
5. Indian Finance Year Book
6. Reserve Bank Bulletin.
7. Wealth of India by Wadia and Joshi.
8. India's National Income by V. K. R. V. Rao.
9. Production in India by R. K. Das.
10. Sankhya.

(B) Foreign.

Publications of Great Britain.

1. Board of Trade Journal and Commerical Gazette.
2. Ministry of Labour Gazette.
3. Journal of Royal Statistical Society.
4. Annual Statistical Abstract of United Kingdom.
5. The Abstract of Labour Statistics.
6. Report of the Controller of Currency.
7. Guide to Current Official Statistics.
8. Statistical Abstract of British Overseas Dominions and Protectorates.
9. Annual Statement of the Trade of the United Kingdom with foreign countries and British Possessions.
10. Reports of Commission on National Debt and Taxation.
11. London and Cambridge Economic Survey.
12. Census Reports.
13. Census of Production Report.

Publications of the League of Nations.

1. Monthly Bulletin of Statistics.
2. International Statistical Year Book.
3. Year Book of Labour Statistics.

4. International Labour Review.
5. Methods of compiling Cost of Living Index Numbers.
6. Memorandum of Public Finance.
7. Memorandum of production and Trade.
8. Memorandum of Currency and Central Banks.
9. Method of conducting Family Budget Enquiries.
10. Methods of Compiling Housing Statistics.
11. Memorandum on International Trade and Balance of payments.
12. Proceedings of the International Conference relating to Economic Statistics.
13. Methods of Statistics of Unemployment.

Publications of other Countries

1. Statistical Abstract of the United States of America.
2. Year Book of United States Department of Agriculture.
3. The World Almanac (New York).
4. Review of Economic Statistics (Harvard Economic Service).
5. The Annalist (U. S. A.).
6. Report of the Conference on Trade Cycles (Carnegie Institution).
7. International Crop Report and Agricultural Statistics (International Institute of Agriculture, Rome).
8. International Year Book of Agricultural Statistics (International Institute of Agriculture, Rome).
9. Monthly Review of Business Statistics (Canada).
10. Financial and Economic Annual of Japan.
11. Journal of the American Statistical Association.

APPENDIX VI

SOME IMPORTANT BOOKS ON STATISTICAL METHOD.

1. Boddington A. L. —Statistics and their application to commerce.
2. Bowley A. L. —Elementary Manual of Statistics.
3. Bowley A. L. —Elements of Statistics.
4. Brinton W. C. —Graphic Methods for presenting facts.
5. Connor L. R. —Statistics.
6. Croxton and Cowden. —Practical Business Statistics.
7. Crum and Patton —Economic Statistics.
8. Davis G. R. —Economic Statistics.
9. Day E. E. —Statistical Analysis.
10. Elderton W. P. —Frequency Curves and Correlation.
11. Fisher Irving —Making of Index Numbers.
12. Fisher R. A. —Statistical Methods for Research Workers.
13. Florence. —Statistical Methods in Economics.
14. Haskell A. C. —How to make use of Graphic Charts.
15. Holman L. J. —Simplified Statistics.
16. Holmes R. I. A. —Statistics for Professional Students.
17. Jones D. C. —A First Course in Statistics.
18. Karsten. K. G. —Chart and Graphs.
19. King W. I. —Elements of Statistical Method.
20. Lovitt and Holtzcwlaw —Statistics.
21. Mills F. C. —Statistical Methods.
22. Mudget B. D. —Statistical Tables and Graphs.
23. Neiswanger —Elementary Statistical Methods.
24. Pearson and Bennet —Statistical Methods.
25. Rhodes E. C. —Elementary Statistical Methods.

26. Richardson. —Business Forecasting
27. Riegel L. —Elements of Business Statistics.
28. Riggleman and Frisbee —Business Statistics.
29. Secrist H. —An Introduction to Statistical Method.
30. Secrist H. —Readings and Problems in Statistical Method.
31. Thurstone. L. —Fundmentals of Statistics.
32. Tippet L. H. C. —Statistics.
33. Westergard H. —Contribution to the History of Statistics.
34. Wheldon H. J. —Business Statistics and Statistical Method.
35. Yule and Kendal —An Introduction to the Theory of Statistics.
36. *Zizek* —Statistical Averages.

ANSWERS

V (12) 2, 8, 7, 6, 5, 3, 6, 3, 6, 5, 8, 1, (13) 2, 4, 7, 2, 10, 5, 10, 11, 3, 7, 5, 3, 1, (14) 1, 17, 11, 13, 7, 7, 5, 7, 5, 7.

VI (4) Average Price Rs. 1–3–9-6, Weighted Average Price Rs. 1–1–1.5 (5) Rs. 156–7–4.2. (6) 46.9 (7) 35.7. (8) 125.1, (9) 61.5. (10) 115.2. (11) 3.6. (12) A. Rs. 12–4–8.6 B. Rs. 16–7–0.19, (15) 67.9. (16) 67.9. (17) Median 22.6, Upper Q 33.9. (18) Median 1333, First Decile 163 (19) Median 2883.4, First Q 1059.8 (28) 22.4 (29) 102.8. (30) 16.6. (31) 62.6 (32) 10.69. (34) G. M. 626, H. M. 60.3 (36) A, 13.45, B, 15.0. (37) Town A. (38) In C. P.

IX (8) 359.2, 374.6, 393.2, 407.2, 409.8, 417.0 425.0 423.4, 443.0, 458.4. (9) 263.8, 2 6.6, 272.7, 277.0. 281.0 285.1 288.7, 292.0, 292.0 297.4, 301.7, 304.8, 307.2, 311.5, (10) 691.0. 717.0, 742.6, 766.4, 789.2, 811.0, 816.0. (11) Five yearly, 52.2, 60.5, 67.9, 71.9, 73.2, 72.4, 71.6, 70.8, 69.9, 70.5, 69.7, 69.2, 68.4, 69.3, 69.2, 70.1, 71.8, Nine-yearly, 61.4. 65.5, 69.1, 71.1. 71.9, 71.6, 70.7, 69.9, 69.2, 69.7, 69.8, 69.7, 70.4. (12) Capital Outlay 504.6, 531.6, 560.4, 588.4, 613.6, 622.4, 629.6, Gross Earnings 75.0, 78.2, 80-2, 81.0. 80.8, 77.6, 74.8.

X. (8) Q Dev. 4. 1, S. D. 4.78 (9) M. D. 15.3. S. D. 18.1 (10) Dispersion in 1921, 27.3 in 188.4, 23.3. (11) M. D. 24.3, S. D. 28.0 Ratio 0.86. (12) More deviations in imports Coef. of dispersion with S. D. for Exports 0.12, Imports 0.13: with M. D. for Exports 0.104, Imports 0.112. (13) Variability more in weight. Coef. of variation of height 5.4, of weight 10.9. (14) Variability more for batsman B. coef. of A, 68.6 of B, 73.5. (16) S. D. 11.8 Semi Int. Q., R. 9.36. (17) Median, 4.8 M. D. 1.7, S. D. 2.2. (18) Mean yield 1455.27, S. D. 824.0, (19) In B by Rs. 19.2. In B· (20) Variability more in Kohinoor. Coef. Tata 0.366, Kohinoor 0.493. (21) S. D. of N. D. 13.6, of C. D. 8.6 of S. D. 10.4 (22) M. D. 2.01, S. D. 2.6. (23) More in 1938, Coef in 1914 9.6, in 1938, 9.7.

XI (5) 0.4, (6) 0.3, (7) Coef. of dis. by S. D. 0.15, Coef. of Skewness by Q. Dev, 0.13 (8) Coef. of skewness 0.41. (9) Coef. of skewness A = −0.32, B, 0.23. (10) A=0.68, B. 0.03.

XII. (8)−0.24 (9) 0.29 (10) 0.46 (11) Area and Price 0.33, Price and Yield 0.003, (12)−0.77 (13) 0.99 (14) 0.93 (15) 0.69 (16) 0.86, (17) 0.79 (18) 0.103, (20) 0.16, (21) 0.71, (22) 0.87, (23) 0.56, (24) 0.05. (25) 0.84 (26) 0.79 (27)−0.51.

XIV. (17) 101·37 99.26 (18) 94.0, 95.78, 99.62, 96.11, 96.48 (19) 102.75 (20) 105.54 (21) 96.94.

INDEX

A

Absolute dispersion 210
Absolute error 32
Accounts relating to sea borne trade. 376
Agricultural and Animal Husbandry 374
Agricultural Statistics 373
Algebraic method 309
Annawari system 375
Annual Report of Irrigation 377
" " of Posts and Telegraphs 377, 378
" " of Railway Board 376
Applied Statistics 14
Approximation 30–32
Area diagram 127
Arithmetic average 71–84
  Advantages 83
  Computation 71
  Moving 190
  Weighted 79
Assumptions 304
Attributes of a good sample 27
Average of Relatives Method 326
Angular diagram 133

B

Band of Belt Chart 172
Bar Charts or diagrams 118
Barometer, Economic 329
Base for Index Number 323
Basic or long Period trend 184
Births and Deaths Rates 81
Bowley–Robertson Committee Report 360
Bracketted method (mode) 90
Bracket Rank Method 284

C

Captions 65
Cartogram 139
Census 25
Census of Production 367
Census Statistics
Chain Base Index Number 324
Chain Base Method 327
Chance selection 29
Characteristics of Statistics 7
Circular diagram 132
Classification 49
Class interval 62
Class limits 63
Coefficient of correlation 269
" " Variation 230
Collection of Data 35
Commerce and Statistics 11
Compensating errors 33
Complex tabulation 56
Composite Average 79
" Units 22
Concurrent deviation method 281
Continuous series 52
Continuous variable 60
Corrected Death rate 81
Correlation coefficient 269
  Direct 257
  Lag 258
  Methods 259
  Negative 257
  Positive 257
Scatter diagrams 266
Cost of Living Index 335
Cost of Living Statistics 381
Crossing Fives 54
Crossing Tens 58
Cumulative error 33
Cumulative frequency table 68
Cumulative frequency curve 154
" " error 215

Cyclical Movements 185
Cubed deviation method 248

D

Data 35
Decile 89
Defects of Government Statistics 3
Degree of accuracy 20
Derivative tables 55
Descriptive Average 82
Descriptive Statistics 14
Diagrams 112-144
Direct personal investigation 36
Discontinuous series 52
  " variable 60
Discrete series 52
Dispersion 209, 244
  Absolute 210
  Relative 210
Distruct of statistics 12
Divisions of statistics 14
Dot Diagram 366
Dun's Index Number 338
Duo-directional diagram 122

E

Economic Adviser's Index No. 361
Economic Barometer 329
Enumerators 39
Errors 32-34
Errors of interpretation 346
Estimates from correspondents 37
Estimates of Area and Yield 373
Extensive sampling 28
Extrapolation 302-315

F

Fallacies
False line 148
Field 57
Fixed Base Index No. 323
Fluctuations
  Irregular 186
  Long-time 184
  [illegible] 105
  Short time 185
Forms of tabulation 55
Free-hand smoothing 187
Frequency, cumulative 68
Frequency curve 154
Frequency curve, cumulative 154
Frequency polygon 153
Frequency tables 60, 65

G

Galton graph 168, 289
Geometric mean 94
Government publications 367, 371
Graph 145-182
Graphic method 259, 306
Grouped data 278
Groups 51
Guide to Current Official Statistics 371

H

Harmonic mean 96
Headings in tabulation 65
Historical series 157, 183
Historigram 183
Horizontal bar charts 124

I

Importance of Statistics
  Index Number 319-342
  Consrtuction 320
  Chain base 225
  Choice of items 321
  Fixed base 328
  Prices 322
  Purpose of 320
  Selection items for 321
  Uses of 328
  Weightage in 323
Index of business Activity 329
Index of Cost of living 240
Index Number of Indian Prices 380
Indian Economic Enquiry Committee 359
Indian Roads 376

Indian Trade Journal 375, 379, 380
Indirect oral investigation 37
Industrial statistics 378
Inertia of large numbers 27
Interpolation 302
Interpretation 343
Investment Index Number 337
Investor's Chronicle 337
Irregular fluctuations 186
Isotype method 137

L

Lag 190, 258
Lagranges' formula 313
Large Industrial Establishments 378
Law of Inertia of Large Numbers 27
Law of Statistical Regularity 26
Lead 258
Least squares 194
Less than diagram 155
Limitations of Statistics 11
Line of best-fit 195
Line diagram 118
Line of equal variation 289
Line of regression 287
Link relatives 119
Logarithmic graphs 110
Logarithms
Long time change 271
Long time Trend 184
Lorenz curve 175, 231
Lower quartile 87

M

Machine sorting 67
Manifold classification 51
Mean
  Arithmetic 71–84
  Geometric 94
  Harmonic 96
Mean Deviation 217
Meaning of Statistics 6
Measuring time changes 187
Median
  Advantages 80
  Definition 84
  Computation 84
  Ogive to locate 86
Method of collectng data 35
Mid-rank method 284
Mode
  Advantages 98
  Definition 89
Modulus 235
Moments of dispersion 235
Monthly Average Corrected for Trend 198
Monthly Statistics of Cotton Spinning 378
Monthly Statistics of production 378
Monthly Survey of Business Conditions 326, 377, 379
Moral and Material progress 359
'More than' diagram 155
Moving Average 190
Moving Average *M*ethod 190
Multiple Bar Chart 122

N

National Income 363
Natural scale graph 145
Negative correlation 257
Nonsense correlation 258

O

Ogive 154
Organization of Statistics 361

P

Pearson's method 268
Percentage distribution diagram 125
Percentiles 89
Periodic variation 185
Pictogram 137
Pin maps 140
Planning an enquiry 19
Plantation statistics 374
Polar diagram 126
Population 25

Position of Averages Method 247
Positive correlation 257
Preliminaries to the Collection of data 18
Price statistics 380
Primary data 35
Primary tables 55
Probable error 269
Progressive Average 193
Projection 303
Punch card 67

Q

Quadratic mean
Quartile deviation 217
  ,, ,, method 247
Quartiles 87
Quequennial Report on Average yield 374
Questionnaries 43-44
Quintiles 89

R

Railway Statistics 376
Random Sampling 29
Range 211
Ratio 22
Ratio of Regression 291
Ratio scale graph 100
Ratio of Variation 286
Regression line 287
Relative 256
Relative measures of dispersion 214, 216, 220
Relatives of Aggregate method 325
Report on the Progress of Broadcasting 378
Report on the Progress of Civil Aviation 377
Representative sampling 28
Rural survey 306

S

Sampling 25
  Method of 28
  Random 29
  Stratified 29
  30
Schedule 268
Scientific Statistics 14
Scope of Statistics 6
Secondary data 35, 45. 46 47
Secular trend 184
S-curve 187
Semi averages method 189
Semi-Inter quartile range 216
Series 51
Series of average
Seriation 52
Short time variation 185, 275
Simple bar diagram 118
Simple classification 51
Simple tabulation 55, 58
Simple unit 22
Single tabulation 55
Skewness 244-255
  Measures of 244
Sources of data 46
Sources of errors 32
Spearman's Method of Rank 283
Spearman's Footrule 286
Spot diagram 55
Square diagram 131
Stability test 80
Stages in statistical investigation 17
Standard Death Rate 81
Standard Deviation 223
Statistical Abstract in India 359, 361 374, 376, 377, 378, 379
Statistical data 50
Statistical Laws 2, 6, 27
Statistical Method 14
Statistical regularity 26
Statistical data 85
Statistics in India 358
Strata chart 175
Stratified sampling 29
Sub-divided bar 124
Surface chart 178
Symbols

T

Tabulation 52, 53, 65
  Mechanical aids in 67
  Stages 53
Tally sheets 54
Ties in ranks 286
Trend 134-187
  Cyclical movements 185
  Long-Time 184
  Seasonal 185
  Secular 184
  Short-time 185
Trade Statistics 382
Transport statistics 375
Triennial Review of Irrigation 377
Types of Statistical unity 22

U

Unbiassed error 33
Undulatory movements 185
Unit of Statistical data 20
Universe 25
University city survey 367
Urban survey 367

V

Value of Statistics 8
Variance 223
Variate 60
Variation 230
Vienna method 137
Volume diagram 135

W

Wages statistics 301
Weightage 323
Weighted mean 79
Working class Index Number 331
Working Sheets 53

Z

Zee Chart 170
Zero line 147, 148
Zero Chart or Curve 172

INSURANCE

PRINCIPLES AND PRACTICE

BY

R. S. SHARMA, M.COM.. LL.B.

Professor of Insurance & Statistics

H. L. College of Commerce Ahmedabad

with a Foreword by

Dr. V. K. R. V. RAO, M. A. D. LIT.

Cr. 16. 458 pages. 2nd Revised Ed. 1953 *Rs. 6/8.*

For a long time a great need was felt for a comprehensive text book on the subject of Insurance written on conditions operating in India and in the background of Indian Insurance Law, Prof. Sharma has achieved this object admirably and he has treated in detail the principles on which various kinds of insurances are based. Relevant portion of the Insurance Act is also given. A very useful book for the students of Insurance as well as for Insurance Companies. Dr. Rao writes in his foreword :—

" Prof. Sharma's work is distinguished for lucid style and clarity of exposition. He has dealt with a complicated subject in a very simple manner and made it not only instructive but also interesting. I am sure it would serve as an excellent text book for the class for which it is intended."

VORA & CO., PUBLISHERS LTD.
3, ROUND BUILDING, KALBADEVI ROAD, BOMBAY.

COMPANY SECRETARIAL PRACTICE

By

M. SATYANARAYANA, M.A. (Com.), B. L.

Professor of Business Organisation and Secretarial Practice. H. L. College of Commerce, Ahmedabad

A standard book now published for the use of students of SECRETARIAL PRACTICE of the Inter Com. examination of the University of Bombay.

Written in full compliance with the syllabus prescribed by the University and with special reference to the law and practice in India.

Contains exhaustive explanations, definitions, forms, notices, resolutions, reports, circular, outlines of each chapter, and subjective index.

Specimen Questions given on each chapter with page reference of answers to the same, for ready reference.

38 chapters—5 Appendices.

Demi 8 48 pages Price: *Rs. 10/8*

Very useful Book of Reference for Company Managers, Secretaries & Commerce Students

VORA & CO., PUBLISHERS LTD.
3, ROUND BUILDING, BOMBAY 2.

AN OUTLINE
of
BANKING SYSTEMS IN INDIA

BY

M. V. SUBBA RAO. M. A., LL. B.

Podar College of Commerce & Economics, Matunga.

This is a book dealing with Indian Banking in all its aspects written in a very simple style and making for the first time this intricate subject understandable to the students of Indian Banking. In simplifying the subject the author has not omitted any important subject connected with Banking but has dealt with everyone of its aspect and also cited enough statistics regarding it. Everything that is connected with Banking including even the Company Law has been treated in this book. The book is up-to-date and will be found very useful to the students of Indian Banking.

Contents

Development of Banking in India—Problem of Royal Credit. Money-lenders and Bankers, Co-operative movements-Long term Finance—Land Mortgage Bank Debts i.e., Condition and Redemption—Commercial Bank-Insurance Companies—Loan Offices, Nidhis, Chit Funds—Industrial Finance—Financing Foreign Trade—Exchange Banks Imperial Bank of India—Commercial Banking—Indian Joint Stock Banks—Reserve Bank of India—Money Bill and Capital Markets—Bank Failures and Banking Legislation, etc.

Demi 8, Pages, 162, Full-cloth. Price Rs 6.

VORA & CO., PUBLISHERS LTD.

3, Round Building, Kalbadevi Road, Bombay 2.

INDIAN COTTON TEXTILE INDUSTRY (DURING 20TH CENTURY)

By Prof. N. H. Thakkar, M. Com., LL. B.

Demi 8 Pages 278 *Rs. 15/-*

The author gives a detailed account of the rise and growth of India's premier industry, the numerous problems which the Industry has faced during two Great World Wars; working of the Controls and the present plight of the Industry together with chapters on Mill Management and the different varieties of cloth. A complete and detailed survey of the Textile Industry for the students of Economics.

FORMS OF ECONOMIC ORGANISATION

or

GENERAL ECONOMIC INDUSTRY UPTO 1914

By S. B. Mahabal. M. A.,
Professor of Economics,
Pratapsinhji College of Commerce, Baroda.

Crown 8 Pages 288 *Rs. 5/12-*

The author has traced the growth of Economic Institutions from the earliest times to the beginning of the first Great World war. The facts are presented in a very simple and lucid style. It covers the syllabus of the B. Com., course of the Bombay University completely.

ECONOMIC PROBLEMS IN INDIAN AGRICULTURE

By Mahesh Chand M. A.

Second revised edition in Press.

A useful book on Agricultural Economics.

VORA & CO., PUBLISHERS LTD.
3, Round Building, Kalbadevi Road, Bombay 2.

MERCANTILE LAW

By

JAMES MENEZES, B. SC., LL M., Advocate O. S.

High Court Bombay.

Second Revised and enlarged edition

The Hon'ble, Mr. JUSTICE BLAGDEN

Special features of the book

Indian Contract Act, Indian Partnership Act, Indian Sale of Goods Act, with full texts, Comment and Case Law.

* * * *

Complete and exhaustive summaries of Indian Companies Act, Negotiable Instruments Act, Insurance Acts, Insolvency Act, Common Carriers, Registration Act, Stamp Act, and Limitation Act, together with case law.

Royal 8, 489 Pages **Price Rs. 12/-**

THE

NEGOTIABLE INSTRUMENTS ACT

BY

JAMES MENEZES, B. SC., LL.M.

Second revised edition

with a foreword by BHULABHAI DESAI

Principles of the law of Negotiable Instruments and the usages relating it in a lucid, clear and simple language. Together with case law and complete with index.

Royal 8, 168 Pages, 2nd Edition Rs. 4/8-

VORA & CO., PUBLISHERS LIMITED.

3. Round Building, Kalbadevi Road, Bombay 2.

Printed by Bhai Malji at Lipika Press. Kurla Road Andheri
Published by M. K. Vora, For Vora & Co.. Publishers. Ltd.
3, Round Building Bombay 2

DELHI POLYTECHNIC
LIBRARY

CLASS NO. 311

BOOK NO. W 47 B

ACCESSION NO. 8716

MGIPC—S5—XVI-17—11-1-49 2,000